DOCUMENT RESUME

ED 323 248 TM 015 463

AUTHOR	Jaeger, Richard M.; Busch, John Christian
TITLE	A Validation and Standard-Setting Study of the General Knowledge and Communication Skills Tests of the National Teacher Examinations. Final Report.
INSTITUTION	North Carolina Univ., Greensboro. Center for Educational Research and Evaluation.
SPONS AGENCY	North Carolina State Board of Education, Raleigh.
PUB DATE	28 Mar 84
NOTE	392p.; Appendices and references are contained in a separate volume. Page 154 is cropped.
PUB TYPE	Reports - Research/Technical (143) -- Tests/Evaluation Instruments (160)
EDRS PRICE	MF01/PC16 Plus Postage.
DESCRIPTORS	*College Faculty; Communication Skills; Education Work Relationship; Elementary School Teachers; Elementary Secondary Education; Higher Education; Knowledge Level; *Licensing Examinations (Professions); Professional Education; Public School Teachers; Secondary School Teachers; *State Programs; State Standards; State Surveys; Teacher Certification; *Teacher Education; Testing Programs; *Test Validity
IDENTIFIERS	Final Reports; *National Teacher Examinations; North Carolina; *Standard Setting

ABSTRACT

In this final report of an 18-month series of projects designed to investigate the validity of the General Knowledge (GK) and Communications Skills (CS) tests of the National Teacher Examinations (NTE) for North Carolina, highlights and major conclusions are summarized and appropriate standards of performance are recommended. Evidence was collected from 100 public school teachers and 206 college and university faculty members concerning their judgments of using the NTE GK and CS tests to select students for North Carolina's teacher education programs. The types of validity considered were: (1) relevance to students' success in a teacher education program; and (2) students' opportunity to acquire the skills and knowledge that these tests measure (instructional validity). The CS test was judged relevant, as were two of the four subtests of the GK test. Instructional validity of the CS test was deemed acceptable, but judgments of the instructional validity of the GK test were mixed. The minimum standard set for the GK test was 631, and the score established for the CS test was 636. With these standards, almost 20% of students tested in April 1983 would have failed the GK test and over 14% would have failed the CS test. The report includes scripts and forms used in standard-setting sessions. There are 51 tables of data. (SLD)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED323248

CENTER FOR EDUCATIONAL RESEARCH AND EVALUATION

A VALIDATION AND STANDARD-SETTING STUDY OF THE GENERAL KNOWLEDGE AND COMMUNICATION SKILLS TESTS OF THE NATIONAL TEACHER EXAMINATIONS

FINAL REPORT

Principal Investigators
Richard M. Jaeger and John Christian Busch
Professional Staff
Helen DeCasper   Marilyn Mueller   James Sullivan

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
☑ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality

- Points of view or opinions stated in this document do not necessarily represent official OERI position or policy

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
JOHN CHRISTIAN BUSCH
RICHARD M. JAEGER
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

The University of North Carolina
at Greensboro

TM015463

BEST COPY AVAILABLE

A VALIDATION AND STANDARD-SETTING STUDY OF THE GENERAL KNOWLEDGE AND COMMUNICATION SKILLS TESTS OF THE NATIONAL TEACHER EXAMINATIONS

FINAL REPORT

Principal Investigators
Richard M. Jaeger and John Christian Busch
Professional Staff
Helen DeCasper Marilyn Mueller James Sullivan

Center for Educational Research and Evaluation
University of North Carolina at Greensboro

28 March 1984

Project Staff

Principal Investigators

Dr. Richard M. Jaeger
Dr. John Christian Busch

Research Assistants

Ms. Helen S. DeCasper
Ms. Annette Miale
Ms. Marilyn R. Mueller
Mr. James D.Sullivan

Administrative Assistant

Kathy Sumner

Computer Programming and Data Analysis

Dr. Robert Muir

Group Leaders for Data Collection Sessions

Dr. Wanda Powers, UNC-Greensboro, School of Education
Dr. Jeutonne Brewer, UNC-Greensboro, Department of English
Dr. Donata Renfrow, Georgia State University
Dr. D. Michelle Irwin, UNC-Greensboro, School of Education
Dr. Gray Bowman, High Point College, Chemistry Department
Dr. David Pratto, UNC-Greensboro, Sociology Department
Dr. James Sherbon, UNC-Greensboro, School of Music
Dr. Nelouise Watkins, Bennett College, Director of Computer Center

Assistants to the Group Leaders

Ms. Helen DeCasper
Mr. James Palmer
Mr. James Pickering
Ms. Patricia Carter
Ms. Annette Miale
Ms. Patricia Bethea
Ms. Carolyn Brewer
Ms. Laura Wyatt

HIGHLIGHTS

- Evidence on the validity of the National Teacher Examinations (NTE) was obtained from 100 North Carolina public school teachers and 206 college and university faculty members.
- These educators provided judgments on the validity of using the NTE General Knowledge and Communication Skills Tests to select students for North Carolina's teacher education programs.
- The same North Carolina educators provided judgments on acceptable standards of student performance on the NTE tests.
- The 206 faculty members who provided judgments were sampled from all North Carolina colleges and universities with approved teacher education programs. Teachers were sampled from throughout the state.
- Historically black colleges and universities were well represented; over 37 percent of participating faculty members taught at these institutions.
- Two types of test validity were considered in this study: (1) The relevance of the abilities measured by the NTE tests to students' success in a teacher education program, and (2) the students' opportunity to acquire the skills and knowledge measured by the tests (this is called "instructional validity").
- These types of validity have been found by the courts to be sufficient to justify initial use of tests in selecting students for education programs.
- The skills and knowledge assessed by the NTE Communication Skills Test were judged to be relevant to success in North Carolina's teacher education programs.

* The skills and knowledge assessed by two of the four subtests in the NTE General Knowledge Test were judged to be clearly relevant to success in North Carolina's teacher education programs.
* Judgments of the relevance of the NTE Mathematics Test were positive but somewhat inconsistent.
* Judgments of the relevance of the NTE Science Test were mixed. This subtest must be labeled marginally relevant, and further study of its validity is essential.
* Judgments of the instructional validity of the NTE Communication Skills Test were generally positive.
* Judgments of the instructional validity of the NTE General Knowledge Test were mixed.
* The NTE Mathematics Test, Science Test, and Literature and Fine Arts Test were judged to have acceptable instructional validity.
* The NTE Social Studies Test was judged to be marginal in instructional validity. Further study of the validity of this test is essential.
* We recommend that the minimum standard of performance on the NTE General Knowledge Test be 631 points on the Educational Testing Service scale.
* We recommend that the minimum standard of performance on the NTE Communication Skills Test be 636 points on the Educational Testing Service scale.
* With these standards, almost 20 percent of North Carolina students tested in April, 1983 would have failed the General Knowledge Test and over 14 percent would have failed the Communication Skills Test.
* Over 24 percent of North Carolina students tested in April, 1983 would have failed at least one of these tests.

PREFACE

This is the final report of an 18-month series of projects designed to investigate the validity of, and recommend appropriate standards of performance on, the General Knowledge and Communication Skills Tests of the National Teacher Examinations. These studies were supported under contracts between the the North Carolina State Board of Education and the Center for Educational Research and Evaluation at the University of North Carolina at Greensboro.

On October 13, 1978, the State Board of Education and the Board of Governors of the University of North Carolina adopted a resolution creating the North Carolina Teacher Quality Assurance Program. One element of that program is a procedure for assuring the quality of students admitted to North Carolina's teacher education programs.

On March 2, 1983, the State Board of Education adopted the General Knowledge and Communication Skills Tests of the National Teacher Examinations as devices for screening applicants to North Carolina's teacher education programs. Adoption of these tests was endorsed by the Board of Governors of the University of North Carolina on March 11, 1983.

These policy actions necessitated a series of studies to determine the validity of the General Knowledge and Communication Skills Tests of the National Teacher Examinations for the purposes intended by the Boards, and the establishment of recommendations on minimum standards of performance for admission to North Carolina's teacher education programs. This report contains a complete description of the procedures and results of a study designed to provide this information.

The findings contained in this report reflect the professional judgment and opinions of the authors, and do not necessarily represent the policies of the State Board of Education or the Board of Governors of the University of North Carolina. Although more than 300 North Carolina educators contributed judgments that underlie our recommendations, their concurrence with the recommendations cannot be assumed.

Many people contributed to this study, and we wish it were feasible to acknowledge each of them. However, we can only express our thanks collectively to the hundreds of North Carolina educators who provided expert judgments on the central questions the study addressed. It is a pleasure to acknowledge the contribution of the following persons, each of whom provided invaluable assistance, without which, this study would have been impossible. Ms. Toni Patterson, Dr. Roger Schurrer and Dr. Robert D. Boyd of the State Department of Public Instruction have supported the project since its inception in 1982. In addition to providing critical information on the structure of teacher education programs in North Carolina, they have kept us informed of State Board of Education policies and essential elements of the North Carolina Teacher Quality Assurance Program. Dr. David H. Reilly, Dean of the School of Education at UNC-Greensboro, has supported the project by graciously making available the facilities and resources of the School, in addition to essential advice and counsel. Ms. Kathleen Sumner assumed the onerous but critical management of all business-related activities, including fiscal accounting and purchasing. We are most grateful for her expertise and contribution throughout the project.

Three doctoral students in the School of Education at UNC-Greensboro, Ms. Helen S. DeCasper, Ms. Marilyn Mueller, and Mr. James D. Sullivan, and one masters student, Ms. Annette Miale, provided outstanding professional contributions to the substance and management of this project. We hope that the project has enriched their education commensurate with their enrichment of our work.

Although she was never compensated from project funds, Ms. Anita Hawkins has helped us immeasurably with countless administrative and procedural tasks. We hope she realizes how much we appreciate her services.

Finally, we want to thank Dr. Leonard Simmons, Superintendent of Schools in Alamance County; Mr. Fred Bishop of the Central United Methodist Church of Asneville, North Carolina; and Maj. Terry Fowler, USAF, of Seymour-Johnson Air Force Base, for coordinating activities at our three data-collection sites.

Each of these persons should rightly assume a portion of the credit, should this project be judged favorably. However, we bear sole responsibility for any shortcomings.

Richard M. Jaeger and John Christian Busch
Greensboro, North Carolina - March, 1984

TABLE OF CONTENTS

Project Staff i

Highlights ii

Preface iv

INTRODUCTION 1

- The Quality Assurance Program 1
- Summary of the Problems Addressed 2
- Organization of the Report 9

DESIGN OF THE STUDY 10

- Overview 10
- Selection of Participants 14
- Procedures for Collection of Data 31
- Data Reduction Procedures 186
- Data Analysis Procedures 213
- Analysis of Standard-Setting Data 224
- Analysis of Participants' Comments on Tests and Procedures 252

RESULTS 283

- Participation in the Study 283
- Test Validity 299
- Instructional Validity Data 317

RECOMMENDED TEST STANDARDS 334

- Results for Individual Subtests 334
- Results for the General Knowledge and Communication Skills Tests 340

CONCLUSIONS AND RECOMMENDATIONS 362

- Use of the NTE in North Carolina 362
- Recommended Test Standards 363

Essential Considerations and Actions 363

APPENDICES 369

APPENDIX A - Participants Who Reviewed the National Teacher Examinations 370

APPENDIX A1 - Participants Who Reviewed the General Knowledge Test: Science Subtest 371

APPENDIX A2 - Participants Who Reviewed the General Knowledge Test: Mathematics Subtest 374

APPENDIX A3 - Participants Who Reviewed the General Knowledge Test: Social Studies Subtest 376

APPENDIX A4 - Participants Who Reviewed the General Knowledge Test: Literature and Fine Arts Subtest 379

APPENDIX A5 - Participants Who Reviewed the Communication Skills Test: Reading Subtest 382

APPENDIX A6 - Participants Who Reviewed the Communication Skills Test: Listening Subtest 384

APPENDIX A7 - Participants Who Reviewed the Communication Skills Test: Writing Multiple-Choice Subtest 387

APPENDIX A8 - Participants Who Reviewed the Communication Skills Test: Writing Essay Subtest 390

APPENDIX B - Participants' Comments on NTE Subtest Items 394

APPENDIX B1 - Participants' Comments on the Science Subtest Items 395

APPENDIX B2 - Participants' Comments on the Mathematics Subtest Items 420

APPENDIX B3 - Participants' Comments on the Social Studies Subtest Items 433

APPENDIX B4 - Participants' Comments on the Literature and Fine Arts Subtest Items 450

APPENDIX B5 - Participants' Comments on the Reading Subtest Items 468

APPENDIX B6 - Participants' Comments on the Listening Subtest Items 485

APPENDIX B7 - Participants' Comments on the Writing Multiple-Choice Subtest Items 506

APPENDIX B8 - Participants' Comments on the Writing Essay Subtest 525

APPENDIX C - Participants' Comments Concerning Evaluation of the Data-Collection Sessions 531

APPENDIX C1 - Participants' Comments Concerning Evaluation of Data-Collection Sessions for the Science Subtest 532

APPENDIX C2 - Participants' Comments Concerning Evaluation of Data-Collection Sessions for the Mathematics Subtest 535

APPENDIX C3 - Participants' Comments Concerning Evaluation of Data-Collection Sessions for the Social Studies Subtest 538

APPENDIX C4 - Participants' Comments Concerning Evaluation of Data-Collection Sessions for the Literature and Fine Arts Subtest 540

APPENDIX C5 - Participants' Comments Concerning Evaluation of Data-Collection Sessions for the Reading Subtest 542

APPENDIX C6 - Participants' Comments Concerning Evaluation of Data-Collection Sessions for the Listening Subtest 545

APPENDIX C7 - Participants' Comments Concerning Evaluation of Data-Collection Sessions for the Writing Multiple-Choice Subtest 548

APPENDIX C8 - Participants' Comments Concerning Evaluation of Data-Collection Sessions for the Writing Essay Subtest 552

APPENDIX D - Breakdown of Invitees/Participants by Subtest and Institution 555

APPENDIX D1 - Number of College and University Faculty Invitees,by Institution and by Subtest 556

APPENDIX D2 - Number of College and University Faculty Agreeing to Participate, by Institution and by Subtest 560

APPENDIX D3 - Number of Participants by Institution and by Subtest 564

REFERENCES 568

INTRODUCTION

The Quality Assurance Program

Recent research (Weaver, 1979) shows that college students who major in education, on average, score well below students in other majors on various measures of scholastic aptitude. In addition, Weaver states that

> "...there is little evidence to suggest that education faculties sort out weak students prior to teaching or graduation."

The 20-point decline (581 to 561) in total National Teacher Examinations (NTE) scores over the years 1969-70 to 1974-75 supports this contention.

A recent Gallup Poll (1979) indicated that 87 percent of parents of public school students believe that, before prospective teachers are hired, they should be required to pass a state board examination to demonstrate their knowledge of the subjects they will teach.

The North Carolina State Department of Public Instruction (SDPI) has coordinated the development of a systematic approach to strengthening teacher preparation and professional certification in North Carolina. On October 13, 1979 the State Board of Education and the Board of Governors of the University of North Carolina pledged themselves and their staff members to a program designed to ensure the quality of profesional personnel recruited for the public schools in North Carolina. The program is called the North Carolina Teacher Quality Assurance Program. One element of the concurrent resolution of these two boards was:

> "The development of more effective screening and counseling procedures to ensure that students who wish to enter professional teacher education programs meet acceptable standards in such academic areas as English and English usage, literature, the fine arts, social studies, mathematics, and science; such procedures are to be administered prior to admission to professional teacher education programs."

Summary of the problems addressed

State Board action on the NTE Core Battery. After considerable study of alternative methods by which the goal to screen candidates could be achieved, on March 2, 1983 the State Board of Education adopted a resolution requiring the use of Core Battery Tests I and II, (the General Knowledge Test and the Communication Skills Test, respectively) of the National Teacher Examinations (NTE) in screening applicants for admission to North Carolina's teacher education programs. More specifically, the Board resolved as follows:

> "Beginning with entering freshmen in Fall, 1983, and thereafter, all prospective teacher education students in the public and private institutions of higher education in North Carolina with approved teacher education programs must successfully complete the NTE on the following schedule for initial certification:
>
> "...All prospective teacher education students who

> are enrolled as sophomores in institutions of higher education programsin Fall, 1983 must successfully complete Core Batteries I and II of the NTE prior to the completion of their junior year to be formally accepted into an approved teacher education program."

The Board of Governors of the University of North Carolina endorsed this resolution on March 11, 1983.

The NTE tests. The NTE General Knowledge and Communication Skills subtests were selected as devices for screening applicants to teacher education programs because they appear to be consistent with the general areas of competence recommended by the Quality Assurance Program's Liaison Committee for teacher education and the concurrent resolution of teh State Board of Education and the Board of Governors of the University of North Carolina. The areas of competence they recommend include skills and knowledge in English and English usage, literature, the fine arts, social studies, mathematics, and science. The NTE General Knowledge Test consists of multiple-choice subtests in Science, Social Studies, Literature and Fine Arts, and Mathematics. The NTE Communications Skills Test consists of multiple-choice subtests in Reading, Writing and Listening. In addition, the Communications Skills subtest contains an essay portion that appears likely to meet the need for a device to assess students' writing skills. The Committee determined that:

> "These areas of learning represent the common information, basic skills, attitudes and ideals which all citizens should possess in order to function as

effective members of society" (Liaison Committee Report, p. 11).

Although the State Board of Education formally specified that the NTE General Knowledge and Communication Skills Tests be used in screening applicants to North Carolina's teacher education programs, the Board recognized the importance of studying the suitability of these tests for their intended purpose. The Core Battery of the National Teacher Examinations is a new assessment device that has recently become available in its final form. Therefore, it has not been previously validated for its potential application in North Carolina. Indeed, the use of any portion of the National Teacher Examinations as a device for screening applicants for teacher education programs appears to be without precedent. The Board's recognition of this need to study the NTE was embodied in a series of contracts granted to the Center for Educational Research and Evaluation of the University of North Carolina at Greensboro, to investigate the validity of, and to recommend appropriate standards of performance for, these NTE tests.

Validity Issues. The use of tests to select professionals and students has a rich legal history. This history suggests that tests may be used for selection as long as they are demonstrably appropriate to the occupation or education program for which candidates are being selected. In addition, testing and selection procedures must be fair and equitable.

Through the courts and through application of federal administrative guidelines, it has been determined that Constitutional guarantees of equal protection and due process apply to persons who are required to complete screening tests for job placement or educational admission. In particular, various Supreme Court cases and the regulations of the Equal Employment Opportunity Commission require that screening tests be subject to systematic study, in order to determine that the tests measure what their developers and users claim they measure. Such studies address the validity of screening tests. For a detailed review of applicable legal considerations, see Research-based Operational Plan (1982).

Based on a review of pertinent legal findings, administrative findings and research results, the Research-based Operational Plan recommended a study of the content validity and the instructional validity of the General Knowledge and Communications Skills tests of the NTE, prior to their use in North Carolina for screening applicants to teacher education programs.

It should be noted that several equally acceptable and meaningful interpretations of test validation are possible (see the Research-based Operational Plan, (1982). One interpretation focuses on the predictive validity of a test by examining the degree of association between test scores and some future criterion performance. We have recommended that the State of North Carolina pursue a study of the predictive validity of the National Teacher Exminations by examining the relationships between scores on the General Knowledge and Communications Skills Tests earned by sophomores, and their later academic performance in teacher training programs. We have also recommended the use of assessments of teacher (in-field) performance as ultimate validation criteria. At the time this study was initiated, a predictive validity study of the NTE could not be completed in North Carolina, because criterion performance measures not were available.

The Research-based Operational Plan specifies two approaches to validation which are important, and that are as legally defensible as predictive validation. The question asked and the methodology used in each approach are as follows:

Relevance -- Judgmental Predictive Validity: Are the skills and knowledge being measured by the tests relevant to success in an approved North Carolina teacher education program. In particular, are the skills and/or knowledge measured by the NTE Communications and General Knowledge Tests relevant to successful performance in a teacher education program. Note that the future performance being used as a judgmental reference is not performance as a teacher but performance in an academic training program in teacher education. In seeking an answer to this question, we secured the judgments of college and university faculty in teacher education and in discipline departments that were a part of general education programs, as well as recently-graduated and currently-practicing public school teachers.

Instructional Validity: Have applicants to teacher education programs had a reasonable opportunity to learn the skills and acquire the knowledge being assessed by the NTE tests? That is, is there a reasonable degree of congruence between the instruction offered in the first two years of general education programs in North Carolina's colleges and universities, and the content of the New Core Battery of the National Teacher Examinations? College and university faculty (but not public school teachers) were asked to evaluate this congruence in a number of ways. The design of the validation procedures will be discussed in Section II of this report.

Standard-setting issues. Standard-setting requires empirically-grounded recommendations on appropriate test standards for screening applicants to teacher education programs in North Carolina. An extensive review of the literature on standard setting is reported in the Research-based Operational Plan. The standard-setting issues addressed in this study are summarized below:

Research by Jaeger (1981) and Popham (1982) indicates that different types of judges are likely to provide differing recommendatons on appropriate test standards. It is therefore important and useful to secure the judgements of as many legitimate constituencies as are feasible and affordable, on appropriate standards for selecting students for North Carolina's teacher education programs. For example, recommendations provided by university-based teacher educators may well differ from those provided by practicing teachers or university faculty in substantive program areas. The State Board of Education should be better able to make an informed decision on appropriate standards in light of recommendations from judges representing several different groups.

Studies by Cross (1982b), Jaeger (1981), and Popham (1982) showed that many judges modify their recommended standards when given opportunities to reconsider their initial judgments in light of normative data, information on the failure rates that are likely to result from their recommendations, and to some degree, information on the standards recommended by their peers. On logical and theoretical grounds, both Shepard (1976) and Jaeger (1976) suggested that judges involved in standard-setting procedures be given opportunities to reconsider, and modify, their initial recommendations. We therefore incorporated an iterative standard-setting procedure which included the provision of normative information on examinees' performance on individual test items and entire test forms.

Research by Brennan and Lockwood (1980) and Cross (1982b) indicated that Angoff's standard-setting method resulted in higher reliability and smaller inter-judge variance in recommended standards than was found for a number of alternative methods. Cross (1982b) also found that a modification of Angoff's method resulted in standards that were relatively consistent across tests, and relatively consistent with judges' subjective expectations of examinee failure rates. An added advantage of Angoff's method is that it was used by Educational Testing Service in three standard-setting and validation studies of the National Teacher Examinations. At least two of these studies have withstood the scrutiny of the federal courts in challenges to the use of the NTE in teacher certification.

Details on the standard-setting procedures used in this study are provided in a latter section of this report.

Organization of the Report

The remainder of this report describes the procedures used in the study, and the results of the standard setting and validation activities conducted for the State Board of Education. We have attempted to provide a detailed and comprehensive report of all of our efforts on behalf of the Board.

Section II describes each component of the study design. It includes a brief overview of the procedures in Part A, followed by a complete description of the procedures used in the selection of participants (Part B). The selection procedures were initially described in the Operational Plan and finally modified and described in the final report for Phase II of the project. Part C describes the training of group leaders and assistants, the sequence and timing of data collection, and includes copies of the scripts and handouts used in all data collection sessions. Part D describes all data reduction, data entry (including computer programs) data editing and file creation activities on microcomputers and the VAX 11/780 computer. Part E describes all phases of data analysis.

Part III of this report summarizes the results of the study. The extent of participation by public school teachers and college and university faculty is described, the results of relevance and instructional validity assessments are summarized and recommended test standards are reported.

Section IV the final section contains our recommendations on use of the NTE Subtests in th North Carolina Quality Assurance Program and a discussion of some policy implications of those recommendations.

DESIGN OF THE STUDY

Overview

This report summarizes the final results of nearly 18 months of study to determine the validity of, and appropriate standards for, the General Knowledge and Communication Skills Tests of the NTE. An extensive operational plan for the study was provided to the State Board of Education in November, 1982. The results of an initial validation and a thoroughly-documented pilot study were provided in June, 1983.

During the period 1 July 1983 through 15 March 1984, the project included five major activities: 1) Analysis of data collected by Educational Testing Service during the April, 1983 statewide administration of the Revised NTE Core Battery in North Carolina. This analysis provided data needed during standard-setting activities, in addition to information that is critical to the establishment of policy on appropriate standards for the NTE in North Carolina. 2) Preparation for the collection of judgment data needed to determine the instructional validity and content validity of the Revised NTE Core Battery and to establish recommended standards on these tests. 3) Collection of judgment data on appropriate test standards and test validity. 4) Reduction, analysis and interpretation of judgment data on appropriate test standards and test validity. 5) Reporting of results to the State Board of Education and the State Department of Public Instruction, in addition to consultation with both agencies, and consultation with the Board of Governors of the University of North Carolina as well. Major elements of these activities, together with deliverable reports, are identified below.

ACTIVITY 1: ANALYSIS OF DATA FROM THE APRIL, 1983 ADMINISTRATION OF THE NTE IN NORTH CAROLINA

Secure data tape from Educational Testing Service
Verify format and contents of tape
Compute cumulative frequency distributions of examinee performance
Compute item difficulty values for the entire sample of examinees and for selected subsets of examinees

ACTIVITY 2: PREPARATION FOR COLLECTION OF JUDGMENT DATA

Recruitment and training of personnel
Verification of all data-collection facilities and logistics
Refinement of data-collection procedures
Refinement of data-collection materials
Final correspondence with participants

ACTIVITY 3: COLLECTION OF JUDGMENT DATA

Transportation of personnel, equipment, and materials to data-collection sites
Collection of data at three locations in North Carolina
Written acknowledgement of contributions of participants
Written acknowledgement of contributions of local personnel

ACTIVITY 4: REDUCTION AND ANALYSIS OF JUDGMENT DATA

Establishment of computer data base
Computer verification and editing of data
Statistical analysis of data

ACTIVITY 5: REPORTING OF RESULTS AND CONSULTATION

Interpretation of analyzed data

Preparation of report containing study procedures, results and findings

Preparation of executive summary, including policy recommendations

Consultation with State Department of Public Instruction on results and findings of the study

Presentation of results and findings to the State Board of Education and the Board of Governors of the University of North Carolina

Our recommendations are based on judgments of the validity of the NTE General Knowledge and Communication Skills test, for their intended uses in North Carolina, obtained from a statewide sample of 306 public school teachers and college and university faculty members. Teachers who are graduates of approved North Carolina teacher education programs were sampled, using accepted scientific procedures, from school systems in each of the three major regions of the state. College and university faculty in schools, colleges and departments of education, as well as faculty in discipline departments, were sampled from all North Carolina higher education institutions with approved teacher education programs. Each of these 306 individuals spent an entire day taking a subtest of the NTE, carefully reviewing each item in that subtest, and following a rigorous judgment procedure through which they provided recommendations on the validity of the subtest, and an appropriate passing score on the subtest.

The recommendations of these 306 North Carolina educators were tabulated and thoroughly analyzed in deriving conclusions on the validity of the NTE Communication Skills and General Knowledge Tests, and in establishing our recommendations on appropriate and defensible standards of performance on the tests.

A detailed description of the procedures used for sampling participants, data collection, data analysis and interpretation is presented in a latter section of this report.

Selection of Participants

The design for sampling participants is described in the Research-Based Operational Plan (1982). It required that three types of participants be selected. Selection of judges was to be completed by using a probability sampling procedure that ensured the unbiased representation of a) state-wide populations of public school teachers who were relatively recent graduates of approved teacher education programs and b) college and university faculty who contributed to the education of beginning teachers.

The design proposed that data be collected in three sites: locations that were convenient to colleges, universities and school systems in the Mountain, Piedmont and Coastal sections of North Carolina. Judges of all three types were to participate in a data-collection session at the site that was closest to their institution.

The data-collection sites were selected so as to secure an effectively representative statewide sample of public school teachers who satisfied the criteria described above, without attempting the impractical task of sampling teachers from every school system in North Carolina: One county that best represented each of the three regions of North Carolina on six demographic variables was selected from each region. A complete description of how the representative county from each region was selected can be found in Volume 2 of the final report of Phase II of this project, June, 1983. The three counties selected as data-collection sites were: Haywood County (Mountain Region), Alamance County (Piedmont Region), and Wayne County (Coastal

Region).

The following sections describe procedures used to select public school teachers and college and university faculty to serve as participants from each region.

1. Public school teachers

Once Haywood County, Alamance County and Wayne County had been identified as data-collection sites, it was necessary to identify school systems from which public school teachers would be sampled. Teachers were sampled from all school systems that were within a 35-mile radius of the centers of counties that had been selected as data-collection sites. Using this criterion, 48 school systems were identified as sources of public-school participants. These school systems are listed, by region, in Table 1.

Initial contact with each selected school system was made through a letter to the school superintendent from Dr. Craig Phillips, State Superintendent of Public Instruction. Dr. Phillips endorsed the project, asked each superintendent to cooperate by allowing teachers to participate as project participants, and asked that superintendents nominate contact persons who could provide information on teachers that was not obtainable from state data files. A project staff member then wrote to each contact person to confirm their assignment to the project. Copies of letters from Dr. Phillips and the project staff member are shown below.

The State Department of Public Instruction (SDPI) was requested to provide sampling frames which listed all teachers employed by selected school systems who met five criteria. The

TABLE 1

School Systems Included in the Sampling Frames for Public School Teachers

Mountain: Test Site, Haywood County

Buncombe County
Asheville City
Haywood County
Henderson County
Hendersonville City
Jackson County
Macon County
Madison County
Swain County
Transylvania County

Piedmont: Test Site, Alamance County

Alamance County
Burlington City
Caswell County
Chatham County
Durham County
Durham City
Granville County
Guilford County
Greensboro City
High Point City
Orange County
Chapel Hill-Carrboro
Person County
Randolph County
Asheboro City
Rockingham County
Eden City
Madison Mayodan City
Reidsville City

Eastern : Test Site, Wayne County

Craven County
Duplin County
Edgecombe County
Tarboro City
Greene County
Johnston County
Jones County
Lenoir County
Kinston City
Nash County
Rocky Mount City

Onslow County
Pitt County
Greenville City
Sampson County
Clinton City
Wayne County
Goldsboro City
Wilson County

Letter to Superintendents

State of North Carolina
Superintendent of Public Instruction
Raleigh 27611

A. CRAIG PHILLIPS
SUPERINTENDENT

I know you are aware of the State Board of Education's effort to improve the quality of teachers entering the profession through the Quality Assurance Program. One aspect of the QAP calls for the screening of applicants prior to entry into teacher education programs in communication skills and in general education. The State Board has adopted the new Core Batteries I and II of the National Teachers Examination for this purpose and has contracted with The University of North Carolina at Greensboro to undertake a study to assess the validity of and set standards for this screening device. I'm writing to ask for your support and assistance in this study.

Faculty in public and private institutions of higher education and public school teachers will be asked to provide professional judgment data in establishing the validity of the NTE Core Batteries. Also, they will provide critical recommendations on minimal passing scores for admission to teacher education programs in North Carolina.

Teachers from your school system have been selected to participate. In the near future, the principal investigators for this study will be contacting your office for assistance in updating lists of teachers who will be randomly sampled and invited to serve as judges. It would be helpful if you would identify an individual in your office who can serve as the contact person for this research study. Please forward the name of this person to Dr. John Christian Busch (School of Education, 103 Forney Building, The University of North Carolina at Greensboro, Greensboro, North Carolina 27312).

I'll say thanks in advance for your cooperation and assistance in this important activity.

Sincerely,

Craig

A. Craig Phillips
State Superintendent of Public Instruction

Letter to Public School System Contact Person

THE UNIVERSITY OF NORTH CAROLINA
AT GREENSBORO

Center for Educational Research and Evaluation
School of Education

June 7, 1983

Dear ________________:

Your name has been given to us as the contact person for the Quality Assurance Program study. As we sample teachers and invite them to participate in the NTE validation study which is scheduled for October 1983, we may have to call your office to obtain information.

Thank you for your willingness to assist us.

Yours truly,

John Christian Busch
Associate Professor
of Education

GREENSBORO, NORTH CAROLINA / 27412-5001
THE UNIVERSITY OF NORTH CAROLINA *is composed of the sixteen public senior institutions in North Carolina*
an equal opportunity employer

Letter to Public School Teachers

State of North Carolina
Superintendent of Public Instruction
Raleigh 27611

A. CRAIG PHILLIPS
SUPERINTENDENT

Dear Colleague:

I'm sure you have read about the Quality Assurance Program which has recently been adopted by the State Board of Education. The purpose of the program is to improve the quality of teachers entering the profession. You are being invited to participate in an important research project associated with validating the NTE Core Batteries in Communication Skills and in General Education. This examination will be used as an admission test for entry into teacher education programs. One of the purposes of the project is to involve classroom teachers and faculty in higher education in recommending minimum standards that should be set.

Your professional judgment and involvement as a panel member will make a significant contribution to the teaching profession in North Carolina. I sincerely hope that you will be able to assist us in this important work.

Sincerely,

A. Craig Phillips
State Superintendent of Public Instruction

ACP:TP:ah

Letter of Invitation to Public School Teachers

THE UNIVERSITY OF NORTH CAROLINA
AT GREENSBORO

June 10, 1983

Center for Educational Research and Evaluation
School of Education

Dear _______________:

I am writing to ask if you will provide one day of professional consultation to a project which is being conducted for the North Carolina State Board of Education. The project is one of several which the State Board has commissioned as part of the North Carolina Teacher Quality Assurance Program. I am enclosing a letter from Dr. A. Craig Phillips who asks for your assistance in this very important activity.

One component of the Quality Assurance Program proposes that the newly revised Core Battery (Communication Skills and General Knowledge Tests) of the National Teacher Examinations be used as a screening device for college sophomores who are applying for admission to teacher education. As yet, the NTE has not been validated for that purpose in North Carolina nor have standards of minimal performance been set. Dr. Richard Jaeger and I have been asked by the State Board of Education to carry out a validation and standard setting study of the NTE. We need your professional assistance.

The study will be conducted during October, 1983 at locations in Alamance County, Haywood County, and Wayne County. Randomly selected, full-time teachers who are graduates of approved teacher education programs in North Carolina and who received their initial certification between Spring 1978 and Spring 1981 have been invited to participate. College and university faculty will also participate. We anticipate that the work session, starting at about 10:00 a.m. will be completed by 4:00 p.m.; lunch will be provided. We will be able to offer a modest honorarium of $45.00.

We feel that this project has significant implications for the teaching profession in North Carolina and hope that you will be able to lend your expertise to this endeavor. Will you please mail the enclosed postcard now to let us know whether you can participate? Any questions should be directed to me (379-5517). We will write later to provide you with more details on the procedures for this study.

Thank you for your assistance in this important activity.

Yours truly,

John Christian Busch
Associate Professor of Education

GREENSBORO, NORTH CAROLINA/27412-5001

THE UNIVERSITY OF NORTH CAROLINA *is composed of the sixteen public senior institutions in North Carolina*

an equal opportunity employer

five criteria identified by the Operational Plan (1982) are:

a) a full-time teacher in a sampled school system;

b) a graduate of one of the 44 approved teacher education programs in North Carolina;

c) currently employed in a sampled school system;

c) a holder of a valid teaching certificate as of 1983; and

d) initially certified as a teacher no less than two years and no more than five years prior to 1983; i.e., initially certified between Spring 1978 and Spring 1981.

Lists provided by SDPI were concatenated, and then separated by region. The number of qualified public school teachers employed by the sampled school systems in each region was determined, and these three totals were used to calculate the proportionate representation of qualified public school teachers in each region. Teachers were sampled by using a stratified random sampling procedure, with region as the stratification variable, and proportional representation by region. The numbers and proportions of teachers sampled from each region were as follows:

Region	Proportion of Qualified Teachers	Number of Teachers Sampled
Mountain	0.14	23
Piedmont	0.41	67
Coastal	0.45	74
	TOTAL	164

The total sample size of 164 teachers was based on the numbers of teachers needed to recommend standards on, and review the validity of, the seven multiple-choice tests that compose the General Knowledge and Communication Skills Tests of the NTE, 30 additional teachers needed to consider the Writing Essay Test of the NTE, and a desired oversampling of five percent.

Sampled teachers were invited to participate in the project through two letters, sent together in June, 1983, with a stamped return postcard. One letter, prepared by a project staff member, described the Quality Assurance Program, asked teachers to serve as professional participants, and asked that enclosed post cards be returned as soon as possible. The second letter, from Superintendent Craig Phillips, endorsed the project and asked that teachers cooperate. Copies of both letters are provided below.

Follow-up letters were sent to non-respondents in August, 1983. Replacements were selected in September for those teachers who did not respond to the follow-up and for those teachers who responded that they would not participate.

2. College and university faculty

According to the Research-Based Operational Plan (1982), two types of college and university participants were to be used in the main study: faculty in discipline departments who teach undergraduate general education courses in subjects assessed by the General Knowledge and Communication Skills Tests of the NTE, and faculty in departments, schools and colleges of education who teach such courses. Sampling frames for public and non-public institution were developed separately, because the types of data that were readily available for these institutions differed substantially.

Data available for the public institutions with approved teacher education programs permitted the development of sampling frames of faculty from computerized listings. Faculty in these public institutions were identified with their primary academic affiliations by codes that are part of the Higher Education General Information System (HEGIS).

Lists of qualified faculty from private colleges and universities had to be developed by hand, using data contained in institutions' catalogs, and through data obtained by telephone from the institutions' offices of academic affairs and/or departmental offices. In all cases, information secured from institutional documents was verified through telephone conversations with appropriate academic officers. Contacts with these academic officers were facilitated by letters from Dr. Donald Stedman of the University of North Carolina's General Administration and letters to the chief academic officers of the institutions from one of the project's co-principal investigators. Copies of these letters are provided below.

Of necessity, the sampling design for selecting college and university faculty was complex (for a detailed discussion see Volume 2 of the final report of Phase II of this project, June, 1983). The sampling design was intended to meet a number of objectives: First, to ensure representativeness, a probability sampling procedure that incorporated random selection of participants at the final stage of sampling was to be used. Second, institutions in each region were to be represented in proportion to the numbers of graduates from their teacher education programs in recent years. Third, institutions that traditionally served minority students were to be sampled at a

Letter to College and University Contact Persons

THE UNIVERSITY OF NORTH CAROLINA

General Administration
P. O. BOX 2688
CHAPEL HILL 27514

DONALD J. STEDMAN
Associate Vice President
for Academic Affairs

TELEPHONE: (919) 962-6981

February 9, 1983

MEMORANDUM

To: Chief Academic Officers, The University of North Carolina
Academic Deans, Private Institutions of Higher Education

From: Donald J. Stedman

In 1979 the State Board of Education and the Board of Governors of The University of North Carolina approved a proposal to establish the North Carolina "Quality Assurance Program." One key feature of this program calls for the screening of applicants to teacher education programs in the areas of communication skills and general knowledge in science, mathematics, social studies, literature and fine arts and language usage. The Revised Core Battery of the National Teacher Examinations is being considered as a possible screening procedure. The Core Battery has just become available in its final form but has not yet been validated for its potential use in North Carolina. Therefore, the State Board of Education has requested that the University of North Carolina at Greensboro conduct a study designed to assess the validity of and set standards for this Core Battery. I am writing to ask for your assistance in the conduct of this study.

Professors from public and private institutions of higher education and public school teachers will be asked to make certain professional judgments about items in the Core Battery to help establish the validity of the National Teacher Examinations. Also, they will be asked to make recommendations on minimum scores required for admission to teacher education programs. Pilot study activities will take place during the spring and summer of 1983. The study will be completed in late 1983 or early 1984.

In the near future, one of the principal investigators for this study, Dr. John C. Busch or Dr. Richard Jaeger, will be contacting you for assistance in developing lists of faculty from which certain persons will be invited to serve as judges. Participation will be voluntary but we hope to get full cooperation of faculty members in this important study.

I hope you will be as helpful as possible.

DJS/cw

Letter to College and University Faculty

THE UNIVERSITY OF NORTH CAROLINA
General Administration
P. O. BOX 2688
CHAPEL HILL 27514

WILLIAM FRIDAY
President

TELEPHONE (919) 962-6981

May 27, 1983

Dear Colleague:

The Board of Governors of The University of North Carolina and the State Board of Education have adopted a proposal establishing the North Carolina "Quality Assurance Program." A key feature of this program calls for the screening of applicants to teacher education programs in the areas of communication skills and general knowledge in science, mathematics, social studies, literature and fine arts. While the Revised Core Battery of the National Teacher Examinations is being considered as a possible screening procedure, this test has just recently become available in its final form and it has not yet been validated for this use in North Carolina. Therefore, the State Board of Education has requested that the University of North Carolina at Greensboro conduct a study designed to assess the validity of, and set standards for the Core Battery. I am writing to ask for your assistance in the conduct of this study.

Professors from public and private institutions of higher education and public school teachers will be asked to make certain professional judgments about tests comprising the Core Battery to help determine the validity of the National Teacher Examinations. Also, they will be asked to make recommendations on minimum scores required for admission to teacher education programs. The study will be completed in late 1983.

Dr. John Christian Busch and Dr. Richard Jaeger, the principal investigators for this study, are writing to ask for one day of your professional consultation. Of course, participation is voluntary; however, I do hope that you will be able to lend your valued assistance to this important task.

Thank you for your assistance.

Cordially,

William Friday

WF/cw

THE UNIVERSITY OF NORTH CAROLINA *is composed of the sixteen public senior institutions in North Carolina*
An Equal Opportunity/Affirmative Action Employer

Letter of Invitation to College and University Faculty

THE UNIVERSITY OF NORTH CAROLINA
AT GREENSBORO

June, 1983

Center for Educational Research and Evaluation
School of Education

Dear Professor ______________:

Those of us in higher education all share a concern for educational quality. We are therefore writing to ask you to provide one day of professional consultation to the North Carolina Teacher "Quality Assurance Program" (QAP). This program has recently been endorsed by the Board of Governors of the University of North Carolina and the State Board of Education. The enclosed letter from President William Friday describes the program.

One aspect of the Quality Assurance Program proposes that the Revised Core Battery (Communication Skills and General Knowledge Tests) of the National Teacher Examinations (NTE) be used as a device for screening college sophomores who are applying for admission to teacher education programs in North Carolina. These tests have sections that are titled Science, Social Studies, Mathematics, Literature/Fine Arts, as well as Reading, Writing, and Listening. As yet, these tests have not been validated for their intended use in North Carolina, nor have appropriate standards of performance been set. We have been asked by the State Board of Education to carry out the first of several studies of these tests. The project will secure the professional judgments of faculty members on the validity of the NTE Core Battery and on appropriate standards. We are inviting your participation in this effort.

We feel that this project will be very important to education in North Carolina. Work sessions will be conducted in the (location)___ on October __, 1983. We will be able to provide a modest honorarium. If you can provide us the benefit of your professional judgment on this date, we will be most appreciative. Please mail the enclosed postcard today and let us know if you can participate.

We will write to you late in the summer to provide more specific information regarding location and activities which are planned. If you have questions, please call us at (919) 379-5517.

Sincerely,

Richard M. Jaeger Professor	John Christian Busch Associate Professor

GREENSBORO, NORTH CAROLINA/27412-5001
THE UNIVERSITY OF NORTH CAROLINA *is composed of the sixteen public senior institutions in North Carolina*
an equal opportunity employer

higher rate than institutions that traditionally served non-minority students.

A stratified random sampling plan was developed to meet these objectives, with the following variables used for stratification: region of the state and "minority" versus "non-minority" status. Although statistical criteria were used to identify the numbers of judges to be sampled from each institution for each of the eight NTE tests, these initial values were rounded and modified so as to meet the requirements of the detailed procedures described below.

a. The total number of teacher-education graduates of each institution was calculated for the 1981-82 and 1982-83 academic years. Totals for institutions that traditionally served minority students were multiplied by a weighting factor of 1.5. The total number of college and university faculty of each type needed to fill all eight subtest review panels was estimated to be 156. The number and proportion of faculty of each type to be sampled from each region were calculated as a weighted fraction of the 156 faculty members of each type needed from throughout the state. The weights used were based on the total numbers of teacher education graduates in the two most recent academic years from each region. Within each region, totals for "minority" institutions were inflated by a factor of 1.5.

b. Target sample sizes of each type of participant for each institution were computed as weighted fractions of the total target number for the state (156). Once again, total numbers of teacher education graduates in the two most recent academic years were used as weights, and an inflation factor of 1.5 was used for "minority" institutions.

a. The numbers of participants of each type to be used on the seven review panels assembled for the multiple-choice subtests that compose the NTE were specified in the Operational Plan (1982). An additional review panel, composed of 30 faculty of each type, was specified for the Writing Essay Test as a result of the pilot study. The number of participants of each type to be sampled from each institution and assigned to each of the eight subtest review panels was determined initially by multiplying the number of participants of each type needed for a panel by the proportion of the 156 judges of that type to be sampled from each institution. The resulting numbers were reviewed and adjusted on the basis of the following criteria:

i. Each of the 44 institutions with approved teacher education programs should have at least one faculty member participating in the study. Initially-determined numbers were sometimes "rounded up" to achieve this goal.

ii. Totals of participants to be sampled from each institution and assigned to review each test were to approximate "target" values. Initial values were inspected to determine the magnitudes of discrepancies from "target" values. Adjustments were made when necessary.

iii. Appropriate representation of participants from "minority" institutions was to be

> realized for each of the eight review panels and each of the three geographic regions. Relevant totals were inspected and adjustments were made when necessary.

We were reasonably successful in realizing all of our sampling objectives and in meeting the criteria specified above.

Faculty listed in each of the 16 sampling frames (one for each subtest and each type of faculty member) who taught at each of the 44 institutions were assigned unique numbers, and a table of random numbers was used to select random samples for each of the eight-subtest panels.

Sampled faculty were sent a letter of invitation to participate in the study in June, 1983 from the co-principal investigators and a letter of support for the project from President William Friday of the University of North Carolina, together with a stamped postcard for response. Copies of these letters are shown below. Follow-up letters were mailed to non-respondents in August, 1983. In September, 1983 non-respondents and invitees who declined to participate were replaced by faculty members in a backup sample. A description of faculty participants will be found in a latter section of this report.

Procedures for Collection of Data

Selection and Training of Group Leaders and Assistants. Each panel was conducted by a senior professional group leader who was supported by an assistant. Experience during the pilot study in April 1983, strongly suggested that individuals with expertise in the substantive area reviewed by the panel be selected as group leaders. Some of the group leaders in the present study had experience in that role, during the pilot study while others had served as members of the pilot study panels. In addition, selection of leaders and assistants was guided by considerations of appropriate balance in sex and race of personnel. The names, panel assigned to, and professional affiliation for each group leader are reported in Table 2. The assistants, also listed in Table 2, were all graduate students in the doctoral or masters programs at UNC-G. All group leaders and assistants participated in two or three training sessions prior to data collection. The first session attended by group leaders and assistants lasted approximately two hours and was devoted to providing background on the QAP policy and the National Teacher Examination. In addition, an overview of the types of activities in standard setting and validation was provided. The purpose of the overview was to provide an initial conceptual understanding of the reasons for, and types of activities.

The second training session attended by all personnel lasted approximately two and one-half hours. At this session scripts, handout materials and transparencies were distributed. Drs. Jaeger and Busch went through the scripts, modeling the procedures and answering questions, and modifying the scripts when necessary and appropriate in response to suggestions of the project personnel. The group leaders and assistants were asked to take the scripts home and to review them one more time, and submit any suggestions before data collection activity.

A third training session was held to familiarize the assistants with a behavioral observation system. This system is described in section -- and was used to verify that the procedures were being executed, as originally designed. That session was conducted by Dr. D. Michelle Irwin who was responsible for the development of the observational system.

Roles of the Group Leaders and Assistants. The responsibility of the group leader was to conduct the panel sessions following the scripts provided; they provided directions, announced times, answered questions, and directed the activities of the assistants. They were responsible for the security of NTE test booklets.

The assistants were responsible for the distribution and collection of all materials, they recorded serial numbers for test booklets, acted as "runners" between the panel and research directors, and collected observational data during the sessions. They generally facilitated the work of the panel.

TABLE 2

Group Leaders and Professional Affiliations
Assistants for each Group Leader

Panel	Group Leader (Group Assistant)	Professional Affiliation
Communication Skills		
Reading	Dr. Wanda Powers (Ms. Helen DeCasper)	UNC-Greensboro School of Education
Writing	Dr. Jeutonne Brewer (Mr. Jim Palmer)	UNC-Greensboro Department of English
Essay	Dr. Donata Renfrow (Mr. Jim Pickering)	Georgia Board of Regents Atlanta, Georgia
Listening	Dr. D. Michelle Irwin (Ms. Pat Carter)	UNC-Greensboro School of Education
General Knowledge		
Science	Dr. Gray Bowman (Ms. Annette Miale)	High Point College Chemistry Department
Social Studies	Dr. David Pratto (Ms. Pat Bethea)	UNC-Greensboro Sociology Department
Literature/ Fine Arts	Dr. Jim Sherbon (Ms. Carolyn Brewer)	UNC-Greensboro School of Music
Mathematics	Dr. Nelouise Watkins (Ms. Laura Wyatt)	Bennett College Director of Computer Center

Sequence and Timing of Data Collection Sessions. Panels were convened in three separate locations during the week of October 23, 1983. They were: Central United Methodist Church, Asheville, NC, October 24; the Alamance County Schools, Central Office, Graham, NC, October 26; and Seymour Johnson Air Force Base, Goldsboro, NC, October 28, 1983.

The sessions were planned to begin at approximately 10:00 a.m. and end in the mid-afternoon. These work hours were selected in order to allow panelists a reasonable amount of time to travel to the site in the morning.

Each panel was assigned to work in an individual room which was equipped with an overhead projector, tape recorders and electrical sources (for the listening test), a blackbord and individual desks or tables on which to write.

The sequence of activities is different for the multiple choice and the essay test and will be described separately below;

a. Multiple Choice Test Activities

The scripts which were used by the group leaders are included in Section 3, which follows, and should be consulted for a detailed description of the procedures.

The overall schedule for the multiple choice tests is contained with the scripts. This schedule was also used as a transparency during the panel presentations. The session began with an introduction of the group leader and assistant as well as self introduction by the panel members. The panels were given background information on the Quality Assurance Program, and the National Teacher Examination's Communication Skills and General Knowledge subtests. The specific activities in which the panel would be engaged, validation and standard setting, were described and the schedule for the rest of the day was outlined.

The next activity consisted of the judges "taking" the test which they were going to review. The purpose of that activity was to familiarize the panelists with the test and to give them a student perspective on taking the test.

Following a break, the next activity consisted of "Item Relevance Judgements". The judges were asked to consider the relevance of the skill or knowledge assessed by each test item to success in the teacher training program with which the judge was most familiar. The judge evaluated each item on a four point scale (4="crucial", 3="important", 2="questionable relevance", 1="irrelevant").

The judges were then given a brief stretch-break and began the standard setting activities. The overall standard setting activity consisted of four major parts: Instruction in standard setting (about 10 minutes), 1st standard setting session (about 35 minutes), then following a lunch break, a mini-test exercise, and finally a second standard setting session.

A detailed description of the standard setting activity may be found in the script which is included in this report. Each judge was asked first to consider a minimally competent person applying for admission to their teacher education program; then after reviewing an individual item to determine the probability of that person answering the question correctly.

Following the first standard setting session and lunch, the panelists were provided with a mini-test execise which was used to instruct the judges in the interpretation of two types of "item probability" statistics The first was the proportion of the total population in the North Carolina state-wide testing program who responded correctly to each item; the second probability was the proportion of individuals getting the item correct, who on the total test obtained between 30% and 50% of the items correctly. This second group was used to reference a "minimally competent" group. In addition, the panelists were instructed in the interpretation of a cumulative percentage curve, which allowed them to determine the percentage of individuals state-wide who would fail the test if the standard were set at a given point.

During the second standard setting session, the standard setting forms from the first session were returned to the panelists including the recommended passing score based on the judges' earlier probability judgments. These passing scores were computed on-site using a SCANTRON Model 1200 and a microcomputer which had been programmed to read and compute the standares for each judge. Also, the judges were given the two types of item probabilities and the cumulative percentage information for the test they were reviewing based on the spring, 1983 state-wide administration. Judges were then asked to reconsider their earlier item judgments, given the information provided, and to complete another set of item judgments. The procedure was different from the first standard setting; judges were encouraged to share their ratings with other panel members in a structured leader-directed discussion before re-rating the items. The leaders were provided with information which summarized the ratings for each item, in order to facilitate the discussion. An attempt was made to obtain opinions on relatively "high" and "low" ratings.

At this point public school teachers completed an evaluation questionnaire and were finished. The university and college faculty were asked to provide information on the instructional validity of the test, for the institution with which they were most familiar. There were three separate types of judgments asked for: Analysis of the Test Blueprint, Global Evaluation, and Item-based Validity Judgements.

The judges were first introduced to the general concept of "Instructional Validity", that students should have had a reasonable opportunity to be exposed to learn what a particular test assesses. In the analysis of the test blueprint, judges were asked to consider each part of the test description (blueprint) supplied by Educational Testing Service and answer the question, "Would 90% of the applicants to the teacher education programs at my institution have had a reasonable opportunity to secure instruction in the skills or knowledge defined by each statement on the test blueprint, prior to application for admission to teacher education." Judges answered the question as either "yes" or "no" for each statement on the test blueprint.

The Global Evaluation consisted of a single four point question regarding the overall similarity of the content at the judge's institution. The item-focused validity asked judges to answer the "90%" question for each item on the test. Judges answered "yes", "no", or "?", I don't know.

Following the instructional validity activity, the university and college faculty completed an evaluation questionnaire and were invited, as were the public school teachers, to discuss the process. Throughout the day's activities, the judges were encouraged to write any comments they had about the test, as a whole or its items on an "Item Comment Sheet".

b. Essay Test Activities

The schedule of activities for the Essay Test is somewhat different than that of the Multiple Choice Test. The script contains the schedule. The script used by the group leader should be consulted for a detailed description of the procedures and materials utilized.

The Introduction Session was essentially the same as that utilized for the multiple choice test panels. The only differences consisted of adapting the discussion of validation and standard setting to an Essay Test rather than the Multiple Choice Tests. During the Test Familiarization Session the judges were not asked to "take" the Essay Test. They were provided with copies of the test, the instructions given to students, the information that students probably would bring to the testing session, information on the general procedures used in scoring, and the scoring scale (a "1" to "6" scale).

The next activity was the First Standard Setting Session. The judges were provided with a set of eleven essays, written during the North Carolina state-wide administration in 1983 and the essay questions used with the associated directions to the test-taker. The eleven essays represented one essay at each possible scale point from 2 to 12 and were arranged in order of decreasing quality. Each essay was labled with the name of a tree. The judge was asked to judge whether each essay "satisfies your definition of a minimum standard of acceptable writing for a student who is just beginning a teacher education program in North Carolina." A second set of essays, similarly configured (selected from the same North Carolina administration) was presented with the same directions.

During the Second Standard Setting Session, the 1st Essay set was redistributed, with the scores assigned to each paper by the ETS rater, and a graph of the cumulative frequency distribution that depicted the performance of teacher education applicants in North Carolina. They were instructed in how to use the graph and in the procedure used by ETS to score the essays. The panelists were asked to reconsider their original "acceptable" or "unacceptable" judgments given the information on the test performance in North Carolina and the ETS assigned scores. The panelists were asked to identify the essays that they would consider acceptable and to identify the lowest ETS assigned score which they found to be acceptable. They also identified the highest ETS assigned score which they found to be unacceptable. The process was followed for both sets of Essays.

Following a lunch break, the Third Standard Setting began. The panelists were again given the Essay set from the first standard setting. The group leader was provided with a frequency distribution of the "acceptable" and "unacceptable" responses to each essay. During this session the judges were encouraged to discuss the reasons for their judgments. Each essay was discussed individually. Following that, the panelists were asked to reconsider the essays in light of the discussion and to provide the ETS-assigned score for the lowest acceptable essay and for the highest unacceptable essay in the same fashion as they did during the second standard setting.

Following a break, the panelists were asked to make content (relevance) and instructional validity judgments. Two questions were asked to assess content validity: first, whether the ETS Essay Test actually measures general writing ability and second, whether the skills assessed by the test are relevant to success in teacher education. The content (relevance) judgments were made first.

Public school teacher-panelists were asked to complete an evaluation questionnaire and were then free to leave.

Following this, the college and university faculty were asked to provide information on the instructional validity of the Essay Test; they answered two questions. The first question asked whether most students at their institution would have had a reasonable opportunity to be instructed in the skills described in the ETS Essay Test Specifications Statement. The second asked about the opportunity for instruction in the skills and knowledge that were _actually_ assessed by the ETS Essay Test. The college and university faculty then completed an evaluation questionnaire and were encouraged to discuss their perceptions and judgments of the data gathering activity.

Instrumentation, Scripts and Handouts. The following section includes copies of scripts used by the group leaders to conduct the validation and standard setting sessions. Separate scripts are included for the multiple choice sessions and the essay test session. Immediately following the scripts are the handout mater-ials (including all response forms) used in data collection.

SCHEDULE FOR MULTIPLE CHOICE TEST

Introduction and Briefing	10:00 - 10:20
Test Taken by Judges	10:20 - 10:50
Break and Snack	10:50 - 11:05
Instruction in Relevance Judgment	11:05 - 11:15
Judges make Relevance Judgments	11:15 - 11:50
Stretch-break (in room)	11:50 - 11:55
Instruction in Standard Setting	11:55 - 12:05
First Standand Setting Session	12:05 - 12:40
Lunch	12:45 - 1:55
Mini-test Exercise	2:00 - 2:25
Second Standard Setting	2:25 - 3:10
Public School Faculty: Debriefing Session on Standard Setting and Relevance Judgments College/University Faculty:Break and Snack	3:10 - 3:25
Instruction in Instructional Validity	3:25 - 3:30
Instructional Validity Judgments	3:30 - 4:10
Final Debriefing Session	4:10 - 4:25

Multiple Choice Test Script
Introductory Session (10:00 - 10:20)

Good morning, I'd like to welcome you to this first session of UNC-Greensboro's Quality Assurance Project.

INTRODUCE YOURSELF, INCLUDING YOUR INSTITUTIONAL AFFILIATION AND YOUR AREA OF SPECIALTY, AND THEN INTRODUCE YOUR ASSISTANT.

Before we get started with some background on today's activities, I'd appreciate it if each of you would take just a moment to introduce yourselves. We'll be doing some work as a group later today, and it will be helpful to know each other. At the same time, we'll be passing a roster around. Please print your name and social security number on the roster.

DESIGNATE A STARTING PERSON AND HAVE THE GROUP INTRODUCE THEMSELVES.

Thank you. I want you to know how much we appreciate your willingness to provide consultation to the North Carolina Quality Assurance Program. This program will substantially change the way we select and certify classroom teachers in North Carolina, and that's likely to have a larger effect on our schools in the next few years than anything else I can imagine. The program demands our scrutiny and contribution.

Some of you might not be familiar with the full scope of the Quality Assurance Program, so I'd like to take a few minutes to review it, to place the component we'll be dealing with today in the context of the full program, and to tell you a bit about the National Teacher Examinations.

As you may know, the Quality Assurance Program was first formalized in a joint resolution of the North Carolina State Board of Education and the Board of Governors of the University of North Carolina in 1978. Those Boards recognized that,

SHOW TRANSPARENCY No.1 ON BOARD RESOLUTION

> the state has the responsibility to provide a strong and appropriate instructional program for all its children

and

> The cornerstone of the instructional program delivery system is the competence of the professional personnel charged with the responsibility of operating our schools.

In their joint resolution, the Boards outlined a procedure for ensuring the quality of personnel who would, in the future, be fully certified as teachers in North Carolina's schools. Three steps were proposed: First, to administer a series of screening tests to those who applied for admission to the state's teacher education programs. The tests were to measure applicants' competence in social studies, mathematics, science, literature and fine arts, as well as a variety of communication skills. Second, to administer a set of screening tests to graduates of the state's teacher education programs, prior to granting them initial certification to teach. These tests were to measure the candidates' professional knowledge. Third and finally, to develop a program for evaluating and assisting beginning teachers during their first three years of teaching. This period was to culminate in a final screening, to determine which beginning teachers would be granted full, renewable certification. You can see that the Quality Assurance Program is both ambitious and complex. We can only hope that its positive effects on public education in North Carolina will warrant the state's tremendous investment in the program, and provide our best professional advice on its design and development.

Our focus today will be solely on the first phase of the Quality Assurance Program: selection of applicants to North Carolina's teacher education programs. Other projects funded by the State Board of Education address the other components of the QAP.

The State Board of Education has tentatively adopted two tests from the revised Core Battery of the National Teacher Examinations for use in screening applicants to teacher education programs. The Board is awaiting the results of this project before confirming its tentative decision. In particular, we are to tell the Board whether the NTE tests are valid for use in selecting teacher education students. And if so, we are to define appropriate passing scores for determining which applicants should be admitted and which should not.

Although we will seek your recommendations on only one subtest of the NTE Core Battery today, the [ ] test, I'd like to take a few moments to tell you about the entire test battery.

DISPLAY TRANSPARENCY No. 2 SHOWING THE EIGHT TESTS
OF THE NTE CORE BATTERY, THE NUMBERS OF ITEMS
FOR THE MULTIPLE-CHOICE TESTS
AND THE TIME LIMIT FOR THE ESSAY TEST

The State Board of Education has proposed to use two tests from the Core Battery for screening applicants to teacher education programs: the General Knowledge Test and the Communication Skills Test. Each of these tests has four subtests. The General Knowledge Test consists of a 25-item mathematics test, a 30-item science test, a 30-item social studies test, and a 35-item literature and fine arts test, all in multiple-choice format.

The publisher of the NTE, Educational Testing Service, computes a single scaled score on the General Knowledge test that depends on an applicant's performance on all four subtests.

The Communication Skills Test consists of a 40-item listening test, a 30-item choice reading test, and a 45-item writing test, all in multiple-choice format, in addition to a 30 minute writing essay test.

As was true for the General Knowledge Test, Educational Testing Service reports a single score on the Communication Skills Test.

I've already mentioned that this project is to result in recommendations to the State Board of Education on the validity of the NTE Core Battery and on appropriate standards of performance on those tests for North Carolina's teacher education applicants. At this point, let's discuss the role we're asking you to play in developing these recommendations.

We will consider only two types of test validity. The first type might be termed "judgmental predictive validity". We are going to ask each of you to look at the items in the [ ] subtest of the NTE and make a judgment on the relevance of the skill each item measures to later success in a teacher education program. The second kind of validity we'll consider is called "instructional validity". We will ask those of you who are faculty members at colleges and universities to think about the general education curricula at your institutions, and tell us whether or not most applicants to your institutions' teacher education programs would have the opportunity to learn what the [ ] test assesses, either as a college prerequisite or during their college studies. Because college and university faculty are in a unique position to know the curricula at their institutions, we will not ask those of you who teach in the public schools to advise us on the instructional validity of the [ ] test.

Let me assure you that we recognize the importance of forms of test validation other than these two, and have recommended to the State Board that this project be the first of several studies of the validity of the NTE. For example, assuming the Board confirms its initial decision to use the NTE in screening teacher education applicants, it will be important to examine the relationship between applicants' NTE scores and their later performances in the state's teacher education programs.

Now let's talk about standard-setting. In addition to advising us on test validity, we will ask each of you to help formulate our recommendation on an appropriate passing score for the [ ] test. In doing so, we will ask you to make judgments about the difficulty of each test item, and to consider the abilities of applicants to the state's teacher education programs. I'll be more specific about that process later in the day.

DISPLAY TRANSPARENCY NO. 3 SHOWING THE SCHEDULE OF
ACTIVITIES FOR THE DAY

Let's take a few minutes to review our schedule for the remainder of the day, and then we can begin our first activity.

First, I am going to ask each of you to take the [ ] test, just as an applicant would, so you'll become familiar with the test items. We'll tell you the right answers when everyone has finished so you can score your own test. Then we'll destroy your answer sheets right in this room so that you can be sure you're the only one who knows your score.

Right after you finish taking the test and learning your score, probably around 11:00 o'clock, we'll take a 15 minute break so you can stretch, relax, and get something to eat and drink. We'll have refreshments set up in Room

__________.

After our break, I'll tell you in greater detail how we'd like you to report your judgments on the relevance of the skills assessed by the [ ] test to success in a teacher education program. Then we'll ask you to make those judgments. That will probably take us close to noon. At that point, we'll take a short stretch break, and ask you to bear with us through one more activity prior to lunch.

After describing the nature of the standard-setting activity and the forms to be used for reporting your judgments, we will ask you to provide your initial recommendations on an appropriate standard for the [ ] test. Our estimates of times for these activities are approximate, but it should be somewhere around 12:40 when we finish our first standard-setting session. By then, we guessed that hunger would be the most powerful human need for most of us, so we scheduled a break for lunch.

Lunch will be served in ______________________. We'll meet back here at 2:00 this afternoon for another standard-setting session. That session will begin with a discussion of some information on the performance of North Carolina students when they took the NTE last April. We'll also discuss ways you might take the information into account when you reconsider your initial recommendations on standards. In fact, that's what we're going to ask you to do during the second standard-setting session. We estimate that this session will end around 3:10. At that point, we will ask those of you who teach in the public schools to share some of your reactions to the [ ] test and your feelings about the activities we've asked you to participate in today. At the same time, we'll ask those of you associated with colleges and universities to take a 15 minute break prior to beginning the instructional validity session at about 3:25.

Describing the methods we'll ask you to use in reporting your instructional validity judgments should take no more than five minutes, and we will begin the instructional validity session immediately after that. We should be ready to begin our final session with those of you from colleges and universities at about ten past four. In that final session, we'll ask you to provide the same kinds of information we sought from public school teachers earlier in the day -- your judgments on the [ ] test and your reactions to our validation and standard-setting procedures.

If our time estimates are reasonable, we should be finished with everything no later than 4:30.

Are there any questions?

ANSWER QUESTIONS AS NEEDED

Test Taken By Judges (10:20 - 10:50)

At this time we'd like to ask you to take the [NAME] test just as an applicant would. [ASSISTANT] is passing out a copy of the [NAME] test, number 2 pencils and an answer sheet. You should also receive a sheet titled "Notes on Items". This sheet is for marking your comments on individual test items. We will collect your sheets at the end of the day and after noting your comments, send them to ETS. At ETS's request [NAME] will record the test serial number for each person.

ASSISTANT PASSES OUT NTE TEST BOOKLET #1(W/O ANSWERS,W/O ITEM STATISTICS); NTE ANSWER SHEET; PENCIL;"NOTES ON ITEMS" FORM;ASSISTANT RECORDSTEST BOOKLET NUMBER ON ROSTER

You should know that any judgments you provide will not be individually identified in our reports and that by participating in this project we recognize that you do not necessarily endorse either the North Carolina Quality Assurance Program or the National Teacher Examinations. However, we will list you as a participant in these judgment sessions when we prepare our reports.

ETS requests that we read the following statement:

SHOW TRANSPARENCY #4

You are requested not to discuss the tests during data collection.

We have distributed to you current sections of NTE tests - this material is secure.

To maintain the security of the tests

- you are requested not to remove them from the data collection site; if you leave the room please leave all test material at your seat

- you have received a pre-assigned unique serial numbered test packet - this enables us to identify the recipient of any missing test packets.

- you are asked to make no notes or transcripts of the test questions. Please do not mark or write on the test booklet

- when you are finished, your test packet must be collected. To maintain the necessary security, the materials distributed to you will need to be collected before you leave the data collection room.

ETS is not conducting this validation study. If, however, you have any comments you wish to make regarding the tests, we will be happy to share them with ETS Program and Test Development staffs. We will distribute some sheets for your suggestions.

THEN SCRIPT CONTINUES

Please be assured that we are not interested in knowing your score on the test. The reason we are asking you to take the test is so you can become familiar with the test and be able to view it as a student taking it would.

DISPLAY TRANSPARENCY #5 (ANSWER SHEET)

The ETS answer sheet should be marked with a number 2 pencil. Normally, the front side of the sheet is used to enter identifying information. However it isn't necessary to fill that in. On the other side there is space for you to record your answers by filling in the "ovals" which correspond to your choices.

In about 25 minutes we'll pass out a booklet with the keyed correct answers marked so you can score your own test. Then after you've scored your own test, we'll tear up your answer sheets.

It is now 10:20, please begin to take the test. We should finish in about 25 minutes.

AFTER 25 MINUTES AT ABOUT 10:45 PLEASE SAY

Is everyone finished? If not please try to finish up in the next few minutes. Please keep your answer sheets and "Notes on Items" sheet at your desk. [NAME] will collect the test booklets.

ASSISTANT COLLECTS NTE TEST BOOKLETS #1
AND DISTRIBUTES TEST BOOKLET #2 (WITH ANSWERS)

We will now distribute another copy of the test booklet. This copy shows the answers that ETS has keyed as correct . Take a few minutes to compare those answers to your own and to see what your test score would have been if you were taking the test as an applicant to a teacher education program.

AFTER FIVE MINUTES

[NAME] will pick up your answer sheets and the test booklets. You may leave the "Notes on Items" sheet at your desk.

ASSISTANT COLLECTS ANSWER SHEET AND OBVIOUSLY
TEARS THEM UP IN FRONT OF PANEL ALSO COLLECT THE
TEST BOOKLET WITH ANSWERS KEYED

Break and Snack (10:50 - 11:05)

We will now take a fifteen-minute break. Coffee, doughnuts and cokes are available in Room [NUMBER]. Please return to this room by 11:05.

Instruction in Item Relevance Judgments (11:05 - 11:15)

[NAME] is distributing two documents: a single copy of the [NAME] subtest of the NTE, and a blue computer readable "Item Relevance/Instructional Validity Judgment Form".

DISTRIBUTE: COPY OF NTE FORM #2 (W/KEYED ANSWERS BUT NO ITEM STATISTICS), BLUE COMPUTER READABLE "ITEM RELEVANCE/ INSTRUCTIONAL VALIDITY" FORM

DISPLAY TRANSPARENCY #6 (JUDGMENT FORM)

Please complete the information requested at the top of the form where the circled #1 is shown. Write your name, the name of your college or school, today's date, the department in which you teach or the subject or level at which you teach and the name of the NTE test which you are reviewing, that's the [TEST] test.

Please use a No 2 pencil for all of your responses. Make your marks heavy and dark. If you need a new pencil please let me know.

Next find section "2" on the response form, please write your social security number in the nine square boxes in that section. Then turn the form lengthwise, and blacken the rectangular bubble that corresponds to each digit of your social security number; darken one bubble in each row.

Next, in the square marked with the circled "3", find the number of the NTE test which is being reviewed. This is test number [NUMBER], for the [NAME] test. Then darken the bubble corresponding to that test number.

Now we would like you to make a single judgment for each test item. The basic question is this: An individual item measures a general skill or complex of skills, or part of a body of knowledge. We are asking that you first read the item and determine what skill or knowledge is being measured. Then, we want you to judge how important that skill is to success in the North Carolina teacher training program with which you are most familiar. (College faculty should think about the teacher education programs at their current institution; public school teachers should consider the institution at which they were initially certified.) Last, you should rate the importance (relevance) of the skill on a four point scale. A definition of points on the scale is printed on the blue judgment form (part 4) and is shown in this transparency.

DISPLAY TRANSPARENCY #7

"Crucial" - of critical importance to success in a teacher education program. These are skills or knowledge which a student would certainly need in order to be successful in the teacher education curriculum with which you are most familiar.

"Important" - of great value to success in a teacher education program. These include skills or knowledge which if present in a student would certainly contribute to a successful experience in teacher education; however it would be possible for a student without some of these skills and knowledge to succeed in teacher education.

"Of Questionable Importance" - these include skills or knowledge which might enhance performance in teacher education but which are not essential to success.

"Irrelevant to Success in Teacher Education" - have no value or effect; these skills or knowledge have no bearing whatsoever on successful performance in teacher education.

Remember, we are asking about the relevance of the skill or knowledge the test item assesses to success in a teacher educa-

__tion program__, not to success as a teacher in the classroom. We agree that the relevance of certain skills and knowledge to performance as a teacher in the classroom is important. However, as important as that issue is, it is not the focus of our current study. Are there any questions regarding the judgments we are asking you to make?

If you think the skill or knowledge represented by an item is crucial, darken the rectangular bubble marked "4". If you believe the skill or knowledge is important, but not crucial, darken the bubble marked "3".

If you perceive the skills or knowledge to be of doubtful or questionable relevance, you should darken the bubble marked "2". Finally, if you believe the skill or knowledge measured by the item is "irrelevant" to success in teacher education, then darken the bubble marked " 1".

One judgment is needed for each item on the test. Please record your judgment on the special answer sheet only after you have carefully considered both the item and all of the choices carefully. You should try to make a judgement for each item. However, if you can not make a judgment for an item you may mark the bubble printed with the question mark. In making your judgments, do not be concerned with how many items you are assigning to each of the relevance categories. Treat each item independently and use your best judgment.

Are there any questions?

PAUSE, AND THEN CONTINUE READING

Judges Make Relevance Judgments (11:15 - 11:50)

We expect you to take about 35 minutes to finish this task. Please begin.

AFTER 35 MINUTES READ THE FOLLOWING

It is now 11:50. Is everyone almost finished? [Please take a few more minutes, if necessary to mark your remaining judgments.] The assistant will collect the test booklet, and the response form.

ASSISTANT COLLECTS TEST BOOKLET AND RESPONSE FORM

NOTES ON ITEMS FORM SHOULD BE LEFT WITH PANELIST

Stretch Break

We have one more task before lunch but before we start that, let's stand up and take a stretch break.

STANDARD SETTING
MULTIPLE-CHOICE TEST SECTIONS

Instructions on Standard-Setting (11:55 - 12:05)

READ THE FOLLOWING:

The purpose of this session is to secure your judgments on appropriate standards for the [ ] section of the National Teacher Examinations. In this session you will be asked to judge the probability that a minimally competent applicant to a teacher education program in North Carolina would give the correct answer to each item in the [ ] section of the [General Knowledge] [Communication Skills] subtest of the NTE. Your judgments will be combined with those of others to produce recommended standards on the entire [General Knowledge] [Communication Skills] subtest.

Although the State Board of Education and the Board of Governors of the University of North Carolina will be making final decisions on the levels of skill and knowledge required of applicants to the State's teacher education programs, they believe these decisions should be based on advice from two sources: practicing teachers who have recently graduated from a North Carolina teacher education program, and college and university faculty who are responsible for educating our beginning teachers.

Establishing standards on screening tests is a difficult task. Even measurement professionals disagree on the best method to use. The procedures we will ask you to use today are improved versions of methods that have been used for setting standards in statewide testing programs throughout the country. These methods produced reasonable and replicable standards when they were carefully applied. We expect them to work equally well in North Carolina.

Later today, we will ask for your views on the standard-setting procedures, the NTE test you reviewed, and the test performance you'd expect of minimally competent teacher education applicants. Until that time, it is essential that you not discuss your judgments with other participants. Only in this way, can we be sure that each participant's recommendations reflect only his or her judgments, independent of the feelings of others.

Before we get started, let's review our schedule for recommending test standards.

DISPLAY TRANSPARENCY SHOWING SCHEDULE No. 3

We're going to begin by asking you to make judgments about each of the items in the [   ] section of the NTE. We estimate that this activity will take about 35 minutes, but we want to be sure that everyone has enough time for careful consideration of each test item. So if more time is needed, we'll take it. Following this judgment session, we'll break for lunch. When we reassemble in this room after lunch, we will introduce several types of data that you might find helpful in formulating your later judgments on appropriate standards. We'll spend about 25 minutes discussing these data and some of the ways they can be interpreted. The data we present in this session will be simulated rather than real, so that our interpretations don't bias your later judgments.

At the beginning of our final standard-setting session, we will give you some data on the actual test performance of applicants to North Carolina's teacher education programs. The data will be in the same form as the simulated data we've just discussed. After considering the real data, you will be asked to make a second set of judgments on the test items you examined just before lunch. We expect this session to take about 45 minutes. But again, the time limit is flexible. Are there any questions at this point?

WHEN THERE ARE NO MORE QUESTIONS,
BEGIN THE FIRST JUDGMENT SESSION.

STANDARD SETTING
MULTIPLE-CHOICE TEST SECTIONS

SCRIPT FOR THE FIRST JUDGMENT SESSION
(12:05 - 12:40)

DISTRIBUTE THE [ ] TEST WITH NO ITEM DATA BUT
WITH THE KEY MARKED, TEST FORM No. 2
DISTRIBUTE A JUDGMENT FORM FOR STANDARD-SETTING,
AND PASS AROUND A BOX OF NO. 2 PENCILS

READ THE FOLLOWING:

[Assistant] is distributing a copy of the [ ] section of the [General Knowledge] [Communication Skills] subtest and a computer-readable response form labeled JUDGMENT FORM FOR STANDARD-SETTING.

DISPLAY TRANSPARENCY No. 9 -- JUDGMENT FORM FOR STANDARD-SETTING

A box of Number 2 pencils is also coming around the room. Take one if you don't already have one. Please use a Number 2 pencil to complete the blanks at the top of the form, labeled YOUR NAME, EMPLOYED AT (COLLEGE or SCHOOL) DATE (Today's date is ________), and the COLLEGE DEPARTMENT or SUBJECT/LEVEL TAUGHT blank, depending on whether you teach at the college level or in the public schools. If you teach at the elementary level, please write the grade level you teach this year in the SUBJECT/LEVEL TAUGHT blank. If you teach at the secondary level, please write the subject-area of your courses. Finally, please write in [ ] for TEST REVIEWED. Following the test section name, please write a "1" and circle it. This will identify your response sheet with the first rating session.

After you've completed the top part of the form, please go to the section of the form that asks for your SOCIAL SECURITY NUMBER. That section is labeled with a circled "2". After writing your social security number in the boxes provided, please turn the form lengthwise to grid your social security number in the spaces provided. When you darken the bubbles in the grid, please make your marks heavy and dark. You'll want to darken one bubble in each of the nine rows. Even though we're asking you to put your name on the response form, none of the data you give us today will be reported individually. You can be sure that individual responses will be held in strict confidence. We need your name and social security number on the response only for data analysis purposes.

When you have finished gridding your social security number, please darken bubble [ ] for the [ ] test, on the grid line in Section 3. The line is right below the statement RECORD TEST NO. ON LINE BELOW.

ALLOW ENOUGH TIME FOR EVERYONE TO COMPLETE
THE TOP OF THE FORM

DISTRIBUTE A COPY OF THE DIRECTIONS HANDOUT

READ THE FOLLOWING:

Please read the following directions with me:

In this session we will ask you to make a judgment about each item in the [Reading], [Mathematics], [Writing], [etc.] section of the NTE. The type of judgment is as follows:

You are to imagine a hypothetical applicant who has the minimum amount of knowledge and skill in the subject-area of the [Reading], [Writing], [etc.] test, necessary to succeed in the North Carolina teacher education program with which you are most familiar. You might think of this hypothetical applicant as a "low C" student, at least in the areas of skill and knowledge assessed by the [ ] section of the NTE.

You are to estimate, for each item, the probability or "chances" that this minimally competent applicant would be able to answer the test item correctly.

SHOW TRANSPARENCY ON PROBABILITY No. 10

Probabilities are expressed as decimal numbers and represent the chances or likelihood that something will occur. A probability can range from .00, meaning "not possible", to 1.00, meaning "certain." A probability of .50 corresponds to a 50% likelihood. If you decided the probability that an applicant could answer a test item correctly was .50, you would be saying that he or she had a 50-50 chance of answering the test item correctly.

Since the NTE [ ] test is scored by counting the number of correct answers and has no penalties for wrong answers, an examinee could get full credit on a test item either by knowing the correct answer or by guessing successfully. So when you estimate probabilities, remember that an applicant could get the right answer either through knowledge or with a lucky guess. An examinee with no knowledge of the correct answer who guessed at random on a five-option multiple-choice item would have a probability equal to 0.20 of getting that item right. On a four-option item, the probability of getting the item right by guessing randomly would be 0.25. Measurement research has shown that some examinees, particularly those who do poorly on multiple-choice tests, do not always guess the answer to an item they do not know, even when it is to their advantage to do so. Therefore, actual probabilities of success due to guessing are likely to be somewhat lower than the values we have just discussed.

For each item on the [ ] test, record your estimate of the probability of a correct answer, on the JUDGMENT FORM FOR STANDARD-SETTING. Completely darken the rectangular bubble that corresponds to the interval containing your estimate. For example, if you believe the probability for an item is 0.10 or less, darken the bubble numbered "1" for that item. The "1" bubble represents all probability estimates from zero through 0.10. The bubble numbered "2" represents all probability estimates from 0.11 through 0.20. And so on. The number in each bubble corresponds to the probability value at the upper end of the range it represents. If, for example, you estimated the probability value for an item to be 0.56, you would darken the bubble numbered "6" for that item, since your estimate of 0.56 would be between the probability values 0.51 and 0.60 represented by bubble number "6." Notice that the boxed section of the RESPONSE SHEET marked number 4, summarizes these directions. Are there any questions about this numbering system?

ALLOW ADEQUATE TIME FOR QUESTIONS

GIVE ADDITIONAL EXAMPLES OF THE NUMBERING SYSTEM AS NEEDED

READ THE FOLLOWING:

Now please use the JUDGMENT FORM FOR STANDARD-SETTING to estimate individual probability values for each test item. Do not be concerned about the number of times you darken bubbles with the same numerical value. Just consider each test item independently. Be sure that you darken only one bubble for each test item, that your marks are firm and dark, and that you keep within the bubbles you select. Use a number 2 pencil to record your ratings. We have lots of Number 2 pencils, so if you need another one, please let me know and I'll bring you one. If you change your mind after you have darkened a bubble, be sure to erase your original mark completely.

Remember that for each item, you are to imagine an applicant who is minimally competent in the subject-area covered by the [Reading], [Writing], [etc.] test. That is, a student who barely has the skills and knowledge in this area necessary to succeed in the North Carolina teacher education program with which you are most familiar. You are then to estimate the probability that this minimally competent applicant could answer the test item correctly. Consider the applicant's likely knowledge and the possibility of a lucky guess. Are there any questions?

ANSWER QUESTIONS AS NECESSARY

When you have finished recording all of your judgments, please check your response form to make certain that you have darkened a single bubble for each item in the test section. Then, please raise your hand and [ ] will collect your test booklet, your response form and the directions. If you'll give [ ] your pencils as well, we'll see that they're sharpened before our next session.

It is ____________ o'clock now. Lunch will be served in _______________ when you're finished. Please go right on to lunch when [ ] has collected all of your materials. We will meet back in this room at 2:00 o'clock for our next session. Please begin now.

MAKE SURE THAT YOU COLLECT ALL TEST MATERIALS AND FORMS

STANDARD SETTING
MULTIPLE-CHOICE TEST SECTIONS

SCRIPT FOR THE SIMULATION EXERCISE (2:00 - 2:25)

In the last session before lunch we asked you to make judgments about the likely test performance of a "minimally competent" applicant to a North Carolina teacher education program. We asked you to predict the performance of such examinees on each of the items on the NTE [ ] test. Of course, such judgments were difficult to make. They required you to imagine the capabilities of a hypothetical person and then predict that person's likely test performance.

In the next session, we will provide you with data on the actual test performance of North Carolina teacher education applicants. We will then ask you to reconsider your initial probability judgments. These data will permit you to make your judgments from a somewhat different perspective.

Right now, we would like to introduce you to several types of test performance data, discuss their interpretation, and then describe ways you might want to use data of these types when you reconsider your initial probability estimates.

The data we will use in this session are fictitious, and will be used solely for purposes of illustration. You will get the real data in our next session.

[ ] is distributing a copy of a three-page NTE test labeled "NTE VALIDATION STUDY, Mini-Test Exercise." The test has ten sample items that ETS prepared to illustrate the Social Studies and Reading sections of the NTE. In

addition, [ ] is distributing a one-page graph titled "NTE EXERCISE."

DISTRIBUTE:
MINI-TEST EXERCISE;
NTE EXERCISE (Graph)

Notice that some percentages appear in the left-hand margin of the test. At this point, I'd like to discuss the percentages and the interpretation of the graph. To simplify the discussion, I'll refer to these data as though they were real.

Each of the items on the mini-test has a percentage figure in parentheses written beside it. For example, (98%) appears beside the first item. This means that 98 percent of the North Carolina teacher education applicants who attempted this test item last spring answered the item correctly. As far as these examinees were concerned, Item 1 was very easy. In contrast, they found Item 2 to be somewhat more difficult. Only 77 percent of all applicants answered it correctly, as you can see from the number in parentheses beside the item. Each percentage in parentheses should be interpreted similarly. The higher the percentage, the easier the test item. If you take a moment to scan the ten-item test, you will see that the applicants found Item 1 to be easiest, and Item 9 to be most difficult. Only a third of all applicants who took the test answered Item 9 correctly.

PAUSE FOR A FEW MOMENTS SO PARTICIPANTS CAN
SCAN THE MINI-TEST

How might you use the percentages we've just discussed when you reconsider your initial judgments on the likely performance of "minimally competent" applicants on the NTE [ ] test? If you do choose to consider these data, you might reason as follows: If 33 percent of all applicants who took the test last spring answered Item 9 correctly, an even smaller percentage of minimally competent applicants would be likely to get this item right. So you might consider the 33 percent (equal to 0.33 on the probability scale) to be an upper limit on the chance that a minimally competent examinee would answer Item 9 correctly. You would expect the actual probability that a minimally competent examinee would answer Item 9 correctly to be somewhat lower than 0.33. You could interpret the rest of the percentages in parentheses in the same way -- as upper limits on the expected performance of a minimally competent applicant. But remember that you are not obligated to use data like these at all when you reconsider your initial judgments.

Are there any questions on the interpretation of these data, or suggestions of other ways they might be used?

ANSWER QUESTIONS AS NECESSARY

Now let's discuss another type of data that might be helpful when you reconsider your earlier recommendations. Notice that another two-digit number appears inside a box to the left of each item on the mini-test. The numbers in boxes are always right below the percentages in parentheses. As before, these numbers represent fictitious data, but we'll discuss them as though they were real.

The numbers in boxes were computed using data from last spring's administration of the NTE. They reflect the test performance of a specially selected portion of North Carolina's teacher education applicants, rather than that of all applicants.

To obtain these numbers, we first computed the total test score of each applicant who took the mini-test. After doing that, we selected the test papers of applicants who got between 30 percent and 50 percent of the mini-test items right. Using only those test papers, we computed the percentage of applicants who answered each of the ten mini-test items correctly. These percentages appear in the boxes beside each item. For example, the 85 in the box beside Item 1 indicates that 85 percent of the applicants who answered between 30 percent and 50 percent of the ten items correctly, answered Item 1 correctly. Only 21 percent of these applicants answered Item 9 correctly, as shown by the "21" in the box beside Item 9.

We computed these percentages in order to give you another perspective on the likely performance of "minimally competent" applicants to North Carolina's teacher education programs. We reasoned as follows: First, we would expect the test performance of this subgroup to be somewhat poorer than that of the entire group of applicants. We reasoned, or perhaps hoped, that most applicants would be more than minimally competent. Second, we thought that applicants whose total test scores were somewhere between 30 percent and 50 percent might be close to exhibiting what some would consider minimally competent performance. _You_ might rightly consider a total test score in the 30 to 50 percent range to be well below, well above, or very close to minimal competence. No one will argue with your judgment. Whatever your feelings on minimally competent test performance, we thought you might find it helpful to know how the 30 to 50 percent group performed on each test item.

You might use these performance data as follows: Suppose you felt that those who scored between 30 percent and 50 percent on the mini-test were _less_ than minimally competent. You could then consider the the numbers in boxes to be _lower limits_ on the probabilities for a minimally competent examinee. For example, you would expect the probability for Item 1 to be _more than_ 0.85. Item 9 has a "21" in its box. You would therefore expect its probability to be _more than_ 0.21.

If you felt that the 30 to 50 percent group was _more_ than minimally competent, you could consider the decimal equivalents of the numbers in boxes to be the _upper limits_ on the likely performance of minimally competent examinees. For example, you would expect the probability that a minimally competent applicant would answer Item 1 correctly to be _less than_ 0.85, and the probability for item 9 to be _less than_ 0.21.

If you thought that the group of applicants who scored between 30 percent and 50 percent on the mini-test did represent minimal competence, you could use the numbers in boxes as your estimates of probabilities for minimally competent applicants.

Once again, I cannot emphasize too strongly that we are providing these data solely for your consideration. We are not suggesting that you use them as the sole basis for your recommendations, nor that you necessarily use them at all. We have given you these data only because our experience suggests that judgments can be formed more readily when some data on actual test performance are available.

Are there any questions on the data in boxes, or their interpretation? Does any one have another suggestion on how these data might be used?

ANSWER QUESTIONS AS NECESSARY

ALLOW TIME FOR PARTICIPANTS' SUGGESTIONS

I would now like to discuss the graph headed "NTE EXERCISE". It contains yet another kind of data on the mini-test performance of North Carolina applicants. Again, these data are fictitious.

DISPLAY OVERHEAD OF NTE EXERCISE (Graph) No. 11

On the bottom of the graph you will find a scale marked "Standard," with numbers that range from one to ten. On the side of the graph, the vertical scale is marked "Percent" and has numbers that range from zero to 100. The curved line shows the percentage of all applicants who would have failed the mini-test, depending on where the standard was set. For example, suppose you had recommended a standard (or passing score) of 6 items right on the mini-test. You would be saying that applicants with total mini-test scores of 6 or more should pass, and that applicants with total test scores of 5 or less should fail. Given your recommended standard, the graph will tell us the percentage of applicants who would have failed the mini-test when it was given in North Carolina last spring. Follow along on the overhead and I'll show you how: First, I'll read across the "Standard" scale until I come to the value 6. Then I'll draw a vertical line up to the curve. Where my line intersects the curve, I'll draw a horizontal line over to the percent scale. You can see that my line intersects the "Percent" scale two small lines above 30, so I'll call the percent 32. Therefore, 32 percent of all applicants to North Carolina's teacher education programs would have failed the mini-test, had the standard been set at 6. Since 100 minus 32 is 68, 68 percent of the applicants would have passed the mini-test had the passing score been set at 6.

Let's examine one other case. Suppose your recommended passing score was 4 items correct. In that case, reading over to 4 on the "Standard" scale, drawing a vertical line up to the curve, and then drawing a horizontal line over to the "Percent" scale at the point of intersection, we find that the horizontal line crosses the "Percent" scale at about 8. That means that 8 percent of the North Carolina teacher education applicants would have failed the mini-test last spring, had the passing score been 4 items correct. Also, and 100 minus 8, or 92 percent, would have passed the test.

Notice that a change in the standard from 6 items correct to 4 items correct results in a dramatic increase in the percentage of applicants who would have passed. It increased from 68 percent to 92 percent. Since the mini-test is short, a small change in the standard produced a large change in the percent who would have passed. Since the NTE [ ] test is much longer than the mini-test, the effect of a small change in its standard will not be as pronounced.

Now how could you use information like this when you reconsider your initial item judgments in the next session? Here are some things you might consider:

At the beginning of the next session we will give you a graph for the NTE [ ] test. It will show you the percentage of North Carolina teacher education applicants who would have failed the [ ] test last fall, had the passing score been set at each of the possible score values from one to [maximum score]. In addition, we will return the answer sheet on which you marked your estimated probabilities this morning. We have run those sheets through a scoring machine, to determine the recommended passing score that is implied by your estimated probabilities. You will find this score marked in blue ink at the top of your answer sheet.

You might want to take your recommended passing score, and use the graph to determine the percentage of applicants who would have failed the [ ] test, had your recommended passing score been used last fall. You would use the graph just as we used the one for the mini-test.

If you find that a surprisingly high percentage of appplicants would have failed -- more than you consider to be a reasonable number -- you might want to lower your estimated probabilities on some or all of the test items. If you find that a surprisingly low percentage of applicants would have failed -- far fewer than you expected -- you might want to reconsider your definition of a "minimally competent" applicant, and adjust your estimated probabilities accordingly. We are not suggesting that there is some "ideal" failure rate, or that you should let the overall failure rate implied by your earlier recommendations be the sole determiner of your estimated probabilities. However, it may be important for you to know the implications of your recommendations, in terms of the percentage of failures that they imply.

I want to empahsize once again, that you do not have to use the graph at all when you reconsider your original recommendations. The data you choose to use or ignore are entirely up to you.

Are there any questions on our discussion of the mini-test exercise graph? Does anyone want to suggest other ways these data might be used?

ALLOW SUFFICIENT TIME FOR QUESTIONS AND SUGGESTIONS

COLLECT ALL MATERIALS - EXERCISE GRAPH, MINI TEST EXERCISE

You may want to stand for a moment and stretch. We will begin our last standard-setting session in a few minutes.

SECOND STANDARD-SETTING SESSION (2:25 - 3:10)

READ THE FOLLOWING:

[ ] is distributing four things: A new copy of the NTE [ ] test, a new JUDGMENT FORM FOR STANDARD-SETTING, a graph, similar to the one we used in the Mini-Test Exercise, and the JUDGMENT FORMS that you completed during the standard-setting session right before lunch. Please raise your hand when [ ] calls your name so that (s)he we can be sure you get your own JUDGMENT FORM.

The [ ] test that [ ] is distributing has two numbers beside each item, just like the mini-test exercise. I'll have more to say about those numbers in a moment.

We're returning the JUDGMENT FORM you completed before lunch so you can review your recommended probabilities for individual test items.

DISTRIBUTE THE FOLLOWING:

TEST NUMBER 3
THE JUDGES' ORIGINAL JUDGMENT FORM FOR STANDARD SETTING
CUMULATIVE DISTRIBUTION GRAPH FOR THE NTE [ ] TEST
A BLANK JUDGMENT FORM FOR STANDARD SETTING

At this time, pleas take a few minutes to fill in the blanks at the top of the new JUDGMENT FORM FOR STANDARD-SETTING, just as you did in our session right before lunch. Following the test section name, please write a "2" and circle it. This will identify your response sheet with the second rating session.

SHOW TRANSPARENCY OF JUDGMENT FORM No. 9

Please write and grid your SOCIAL SECURITY NUMBER in the section marked "2", and then darken bubble number [ ] in the section marked "3", to indicate that you are reviewing the [ ] test, just as you did in the session before lunch.

Are there any questions?

ANSWER QUESTIONS AS NECESSARY

Remember, none of the data you provide will be reported individually. We need this information only for record-keeping purposes during data analysis. Your individual responses will be held in strict confidence. Please complete the top of the form now.

PAUSE TO ALLOW PARTICIPANTS TO COMPLETE TOP OF FORM

During this final standard-setting session, we are giving you the chance to reconsider your earlier judgments, in light of data on applicants' actual performance on the NTE [ ] test.

The recommended passing score that we computed from your earlier probability judgments has been written in blue in the upper right-hand corner of the JUDGMENT FORM we have returned to you. Is there anyone who doesn't have a number like that on their form?

SECURE DATA FROM THE ASSISTANT, AS NECESSARY

DISPLAY TRANSPARENCY 11a

Please use this number, together with the graph you just received, to determine the percentage of North Carolina applicants who would have failed the NTE [   ] test if the standard computed from your recommendations been adopted last spring. Just use the procedure we discussed during the Mini-Test Exercise. Let me know if you'd like help in determining that value.

ALLOW SEVERAL MINUTES FOR READING THE GRAPH

Next, please note the numbers to the left of each item on the [   ] test. The format of these data is the same as in the Mini-Test Exercise. A percentage in parentheses indicates the percent of _all_ applicants to North Carolina teacher education programs who answered the test item correctly. A number inside a box also indicates the percentage of applicants who answered the test item correctly, but _only_ applies to those applicants who gave correct answers to between 30 and 50 percent of the items on the [   ] test. These numbers are based on the test performance of a statewide _sample_ of teacher education applicants when the test was administered last spring.

You may wish to think about these percentages when you reconsider your earlier probability judgments for each of the test items.

Are there any questions?

ANSWER QUESTIONS AS NECESSARY

The procedure we will use in the remainder of this session will be somewhat different from the one we used in our standard-setting session just before lunch. In this session we would like to give you the opportunity to discuss the probabilities you estimated for each item during that earlier session.

It is important to realize that, for most test items, this group is likely to have a wide range of estimated probabilities. This is to be expected. Each of you has your own good reasons for your judgments. The purpose of this discussion is to discover some of those reasons so that all might benefit from your thinking.

GROUP LEADER: SEE PAGE 41 FOR GROUPINGS OF ITEMS
BEFORE BEGINNING THIS SECTION

First we will ask you to review the probability values you estimated for Items [first group of items] during the first standard-settings session. We will ask you to discuss your own choices and your reasons for them. Finally, we'll ask you to reconsider your initial judgments in light of the statistical information you've received and your discussion with your colleagues.

Remember that you are to imagine an applicant who has the minimum amount of knowledge and skill in the subject area covered by the [ ] test, necessary to succeed in the teacher education program with which you are most familiar. You are then to estimate the probability that this minimally competent applicant would be able to answer each test item correctly. We'll ask you to record your probability estimates on the JUDGMENT FORM by darkening one bubble for each item, just as you did in our earlier session.

Are there any questions? If not then let's begin. First please locate Items [first group of items] on the test booklet and the probabilities you estimated during our first standard-setting session. These are marked on the JUDGMENT FORM FOR STANDARD-SETTING we returned to you a few moments ago.

(GROUP LEADERS: WHEN YOU ARE REVIEWING THIS SCRIPT, NOTICE THAT AN EXAMPLE OF THE PRINTOUT SHOWING EACH JUDGE'S RECOMMENDED STANDARD AND THE DISTRIBUTION OF RECOMMENDED PROBABILITIES ON EACH ITEM IS CONTAINED IN YOUR PACKET OF HANDOUT MATERIALS)

Let's take a moment to discuss your estimated probabilities for these items. Is there anyone who would be willing to share their judgments with the group? Did any of you recommend a "high" probability value? say greater than 6, on the 1 to 10 scale for Item 1?

ASK A VOLUNTEER TO DESCRIBE WHY HE OR SHE RECOMMENDED A HIGH PROBABILITY VALUE.

NOTE: YOU WILL KNOW THE DISTRIBUTION OF RECOMMENDED PROBABILITY VALUES FOR EACH ITEM. THE PARTICIPANTS WILL NOT HAVE THESE DATA. STRONGLY ENCOURAGE DISCUSSION IF NO ONE VOLUNTEERS. REVEAL THAT YOU KNOW THERE WERE SOME HIGH AND SOME LOW RECOMMENDATIONS, UNLESS THE DISTRIBUTION SHOWS THAT ALL RECOMMENDATIONS WERE AT ONE END OF THE SCALE.

TAILOR YOUR REQUESTS TO THE DISTRIBUTIONS.

SEEK DISCUSSION ON A GROUP OF ITEMS IN SEQUENCE BEFORE ASKING PEOPLE TO RECONSIDER THEIR INITIAL JUDGMENTS AND MARK THEIR RECOMMENDATIONS FOR THAT GROUP. USE THE TABLE ON PAGE 41 TO DEFINE EACH GROUP OF ITEMS.

Some of you may well have estimated somewhat lower probability values. Did anyone darken bubbles that were, say, less than 4 on the 1 to 10 scale?

ASK A VOLUNTEER TO DISCUSS WHY HE OR SHE RECOMMENDED A LOW PROBABILITY VALUE FOR ITEM 1

ENCOURAGE DISCUSSION AS BEFORE

CONTINUE IN THIS WAY, FOR EACH ITEM IN THE FIRST GROUP. ASK FOR DISCUSSION OF HIGH PROBABILITY VALUES FIRST, THEN LOW PROBABILITY VALUES FOR EACH ITEM.

READ THE FOLLOWING:

At this point, I would like you to reconsider your initial recommendations on items [first set]. In doing so, please think about the discussion we've just had, the percentage values printed beside each item, and the graph you worked with a few minutes ago.

Please mark your estimates on the response form by darkening the appropriate bubble for each of Items [ ] [and] [through] [ ]. The 1 to 10 scale is defined in the handout marked PROBABILITY SCALE, if you want to refresh your memory. Does anyone need a sharp Number 2 pencil?

Remember that you can choose to use or ignore any of the information we've given you or any of the points that were raised in our discussion. We want your judgments on probabilities. Please begin now.

PAUSE

Now let's consider Items [second group]. Let's take a moment to discuss your estimated probabilities for these items. Is there anyone who would be willing to share their recommendation on Item [first item of the second group] with the group? Did anyone recommend a "high" probability, say greater than 6, on the 1 to 10 scale for Item [first item of the second group]?

ENCOURAGE DISCUSSION AS BEFORE

AT THIS POINT, CONTINUE AS YOU DID WITH THE FIRST GROUP OF ITEMS. ASK FOR DISCUSSION OF A LOW PROBABILITY ON THE FIRST ITEM AND THEN ALTERNATE BETWEEN DISCUSSIONS OF HIGH AND LOW PROBABILITIES FOR EACH ITEM IN THE SECOND GROUP.

THEN HAVE PARTICIPANTS RECONSIDER THEIR INITIAL JUDGMENTS ON ITEMS IN THE SECOND GROUP BY MARKING THEIR RESPONSE FORMS

CONTINUE THROUGH EACH GROUP OF ITEMS AS SHOWN IN YOUR ITEM-GROUPING TABLE

ITEMS HAVE BEEN GROUPED BECAUSE THEY FORM A LOGICAL SET, OR BECAUSE OF CONVENIENCE.

ITEM GROUPING TABLE
FOR USE DURING STANDARD-SETTING DISCUSSION
(SECOND STANDARD-SETTING SESSTION)

READING	WRITING	LISTENING	SOC. STUDIES	SCIENCE	MATH	LIT./F.A.
1-2	1-3	1-3	1-3	1-2	1-2	1-2
3-4-5	4-6	4-6	4-5	3-4	3-4	3-4-5
6-7	7-8-9	7-8-9	6-7-8	5-6	5-6	6-7-8
8-9-10	10-11-12	10-11-12	9-10	7-8	7-8	9-10
11-12	13-14-15	13-14-15	11-12-13	9-10	9-10	11-12-13
13	16-17-18	16-17-18	14-15-16	11-12	11-12	14-15
14-15	19-20-21	19-20-21	17-18-19	13-14	13-14	16-17
16	[22-23-	22-23	20-21	15-16	15-16	18-19-20
[17-18-	24-25]	24-25	22-23	17-18	17-18	21-22
19-20-	26-27-28	26-27	24-25	19-20	19-20	23-24-25
21-22]	29-30-31	28-29	26-27	21-22	21-22	26-27
23	32-33-34	30-31	28-29-30	23-24	23-24-25	28-29
24-25	35-36-37	32-33-34		25-26		30-31-32
26	38-39-40	35-36-37		27-28		33
27-28	41-42-43	38-39-40		29-30		34-35
29	44-45					
30						

WHEN THE GROUP HAS COMPLETED THE DISCUSSION OF EACH SET OF ITEMS, GIVE THE JUDGES A FEW MOMENTS TO MAKE SURE THAT THEY HAVE DARKENED A BUBBLE FOR EACH ITEM ON THE TEST.

ASK THE JUDGES TO CHECK THEIR RESPONSE FORMS CAREFULLY

READ THE FOLLOWING:

Has everyone finished marking their recommendations for the last set of items? If so, [   ] will collect all of your materials. Please hand [   ] your test booklet, your JUDGMENT FORMS FOR STANDARD SETTING from Session 1 and from this session, the graph that we distributed earlier, and the probability scale handout.

ASSISTANT SHOULD COLLECT:
NTE TEST BOOKLET, TWO JUDGMENT FORMS,
NTE GRAPH and PROBABILITY HANDOUT

MAKE SURE THAT YOU GET ALL MATERIALS FROM EVERYONE

SCRIPT FOR EVALUATION OF THE STANDARD-SETTING PROCEDURE
FINAL STANDARD-SETTING SESSION (3:10 - 3:25)

READ THE FOLLOWING:

At this time, I'd like to suggest that those of you who are college and university faculty take about a 15-minute break before our session on instructional validity. You'll find refreshments in Room ________. Please return to this room at 3:30. At the same time, I'd like to ask those of you who teach in the public schools to take a short stretch break right in this room, and then to take a few moments to fill out a payment form so we can send you your honoraria. We'll also ask you to complete a short evaluation questionnaire and to share your views about the NTE [   ] test and the procedures we've used today.

ALLOW A FEW MINUTES FOR TEACHERS TO STAND AND STRETCH

DISTRIBUTE PAYMENT FORM, INFORMATION FORM
EVALUATION QUESTIONNAIRE, AND RESPONSE FORM

USE THE FOLLOWING SCRIPT ONLY WITH PUBLIC SCHOOL TEACHERS

Please take a few moments to complete the information form and theblue payment forms so we can send you your honoraria. We must ask that you complete the top portion of two copies of the blue payment form, and that you sign each form. Then, please use the green, computer-readable PARTICIPANT'S EVALUAION RESPONSE FORM to provide your responses to the questions on the PARTICIPANT'S EVALUATION QUESTIONNAIRE. Please be sure to write your name and your social security number at the top of the RESPONSE FORM. Here again, we will respect the confidentiality of your judgments, and use your responses only for data analysis purposes. We will not identify your individual responses in any of our reports.

ALLOW TIME FOR COMPLETING FORMS

WHEN IT APPEARS THAT EVERYONE HAS FINISHED,
READ THE FOLLOWING:

Has everyone completed an evaluation form, the top portion of two blue payment forms, and the PARTICIPANT'S EVALUATION RESPONSE FORM? Make sure you have signed both copies of the blue payment form. [ ] will collect your forms.

This completes our formal standard-setting and validation activities. Do any of you have any comments on either the NTE [ ] test or on the procedures we've used today? I've asked [ ] to record your comments so that we can either pass them on to ETS or consider them ourselves when we redesign our procedures.

ALLOW TIME FOR DISCUSSION

I want to thank each of you for participating in this Quality Assurance Project. Without your help, we couldn't give the State Board sound advice on use of the NTE in North Carolina. It has been a pleasure working with you. Drive carefully on your way home.

Instruction in Instructional Validity (3:25 - 3:30)

As a college or university faculty member, you are in a unique position to advise the State Board of Education on an important issue concerning the validity of the NTE [ ] test.

One concern that has been raised in the courts with respect to the use of screening procedures (including those that use standardized tests) has come to be called "instructional validity." The courts have held that when screening procedures are used they must be "fair." If the Quality Assurance Program is implemented as it is currently planned, it says that students will have to pass a test in [NAME] before they may be admitted to a teacher education program. For that to be a fair requirement we must be sure that those applicants have had a reasonable educational opportunity to master the skills or to acquire the knowledge assessed by the [ ] test. That is, the opportunity to learn these skills must be reasonably available for most students.

We will ask you to tell us whether or not most applicants to teacher education programs at your institution will have had the opportunity to learn what the [NAME] test assesses, either during their college studies or as a college prerequisite. In making that judgment we'll ask you to use three different perspectives on the test. Those perspectives will range from the very global (looking at the test as a whole) to the molecular (examining individual items on the test).

Instructional Validity Procedures (3:30 - 4:10)

Analysis of Test Blueprint (3:30 - 3:45)

In this first set of judgments you will be asked to read a written description of the [NAME] test. The summary, called a "test blueprint", describes the skills or knowledge which should be represented on the test. [ ] is now distributing a copy of the test blueprint judgment form for the [NAME] test.

DISTRIBUTE TEST BLUEPRINT JUDGMENT FORM

First please write your name and institution in the space provided at the top of the form. Next you should verify that your "test blueprint" lists [NUMBER] separate descriptions for different test components.

SHOW TRANSPARENCY #12 OF BLUEPRINT JUDGMENT FORM

We would like you to read each part of the test description and then answer the question which is displayed on the screen behind me for each part:

"Would 90% of the applicants to the teacher education programs at my institution have had a reasonable opportunity to secure instruction in the skills or knowledge defined by each statement on the test blueprint prior to application for admission to teacher education?"

A number of points should be stressed before beginning this activity. First, we are asking about the opportunity to acquire knowledge or skills through prior educational experiences, that is through courses or other instruction. Whether an individual student actually learns the skills is not the main issue. What we want to focus on is whether students have had a reasonable opportunity, through instruction, to acquire the skills or knowledge that are measured by the [NAME] test. Second, you should note that the point at which these opportunities have been made available includes any time up to a student's application for admission to teacher education. While many instructional opportunities may have occurred during an applicant's general education program in the first two years of college, others may have presented themselves at an earlier point in the student's career --for example in high school.

You should answer the "opportunity" question for each statement on the test blueprint either with a "Yes" or "No" response by circling your choice on the judgment form . When you are finished marking your response to statement number "1" on the blueprint you should move on to statement number "2" and so on. Each of your responses should be independent of the all the others. That is, don't worry about how many "yes" or "no" responses you have marked.

Please try to make judgments for each statement on the test blueprint. For the (NAME) test we are evaluating, that means a total of [NUMBER] judgments.

Are there any questions? This activity should probably take no more than 15 minutes. Please begin.

ANSWER ANY QUESTIONS THEN PANELISTS SHOULD BEGIN WORK

Instructional Validity

Global Evaluation (3:45 - 3:50)

Is everyone finished? [If not please take a few minutes to finish your responses.] [NAME] will collect your Instructional Validity Response Sheets and distribute a "Global Evaluation of Instructional Validity" Form.

ASSISTANT SHOULD COLLECT TEST BLUEPRINT RESPONSE FORM,

DISTRIBUTE GLOBAL EVALUATION of INSTRUCTIONAL VALIDITY FORM

First, please write your name and the test name in the space provided at the top of the form. You can see that the question on this form is another way of asking about the instructional validity of the [NAME] test. It asks you to indicate, on a four point scale the degree of similarity between the content of the [NAME] test and the general education curriculum at your institution. Please check the response that best describes your judgment. Let me know when you have finished.

COLLECT GLOBAL EVALUATION FORM

WHEN ACTIVITY IS COMPLETE GROUP LEADER SHOULD BEGIN THE INSTRUCTIONAL VALIDITY (ITEM) FOCUSED ACTIVITY

Instructional Validity

Item Focused Activity (3:50 - 4:10)

RE-DISTRIBUTE TEST BOOKLETS #2 (WITH CORRECT ANSWERS, AND NO ITEM P DATA), BLUE COMPUTER ITEM RESPONSE FORM FOR INSTRUCTIONAL VALIDITY, USED IN ITEM RELEVANCE JUDGMENTS EARLIER

As in the earlier activities in this session, we are now going to ask you to judge whether students have had a reasonable opportunity to acquire the skills and knowledge assessed by the test. However, our focus will be slightly different. We have re-distributed your copy of the [NAME] test and the computer readable Item Relevance/Instructional Validity form you used earlier. Please check to make sure you have the validity form with your name on it.

DISPLAY TRANSPARENCY #6 ITEMRELEVANCE/INSTRUCTIONAL VALIDITY FORM

In this activity we would like you to go through the test item by item and mark your answers to this question: "Would 90% or more of the applicants to the teacher education programs at my institution have had a reasonable opportunity to secure instruction in the skills or knowledge represented by this test item, prior to applying for admission. You should answer "yes" or "no" on the Item Response Form for each item on the test, by filling-in the appropriate bubble in Section 5 of the answer sheet. If your answer is "yes", darken-in the bubble marked "Y"; if "no", darken-in the bubble marked "N". If you can not make a judgment, then you may darken the bubble marked with a question mark "?".

We estimate that this task should take about 20 minutes. Please begin now.

OBSERVE JUDGES, IF WORK COMPLETE BEFORE 20 MINUTES THEN CONTINUE BELOW; OTHERWISE ASK PARTICIPANTS TO FINISH UP. ASSISTANT PICKS UP FORMS AND TEST BOOKLET #2.

Evaluation and Debriefing (4:10 - 4:25)

DISTRIBUTE EVALUATION QUESTIONAIRE, COMPUTER-SCORABLE EVALUATION RESPONSE FORM, INFORMATION FORM, AND PAYMENT FORM.

Our last activity will take only a short time. I'm distributing a form titled "Evaluation Form" which contains several short questions. Basically we are interested in obtaining your reactions to the process which was just used to set standards and assess validity for the [NAME] test. There are just a few questions. There is also an Information Form and a Payment Form to be completed.

Please take a few moments to complete the information form and some payment forms so we can send you your honoraria. We must ask those of you affiliated with non-public institutions to complete the top portion of two copies of a blue payment form, and that you sign each form. We will be able to secure payment information from the Information Form for those of you who are affiliated with the University of North Carolina.

Please use the green, computer-readable PARTICIPANT'S EVALUATION RESPONSE FORM to provide your responses to the questions on the PARTICIPANT'S EVALUATION QUESTIONNAIRE. Please be sure to write your name and your social security number at the top of the RESPONSE FORM. Here again, we will respect the confidentiality of your judgments, and use your responses only for data analysis purposes. We will not identify your individual responses in any of our reports.

ALLOW TIME FOR COMPLETING FORMS

WHEN IT APPEARS THAT EVERYONE HAS FINISHED,

READ THE FOLLOWING:

Has everyone completed an evaluation form, an INFORMATION FORM, PARTICIPANT'S EVALUATION RESPONSE FORM? Those of you not at a UNC campus should also make sure you have completed and signed both copies of the blue payment form. [ ] will collect your forms.

This completes our formal standard-setting and validation activities. Do any of you have any comments on either the NTE [ ] test or on the procedures we've used today? I've asked [ ] to record your comments so that we can either pass them on to ETS or consider them ourselves when we redesign our procedures.

ALLOW TIME FOR DISCUSSION

I want to thank each of you for participating in this Quality Assurance Project. Without your help, we couldn't give the State Board sound advice on use of the NTE in North Carolina. It has been a pleasure working with you. Drive carefully on your way home.

ESSAY TEST SCRIPT
SCHEDULE FOR ESSAY SESSION

Introduction	10:00 - 10:20
Test Familiarization	10:20 - 10:35
Standard Setting Instruction	10:35 - 10:45
First Standard Setting Session	10:45 - 11:25
Break and Snack	11:25 - 11:40
Instruction for Second Standard-Setting	11:40 - 11:50
Second Standard-Setting Session	11:50 - 12:05
Instruction for Third Standard-Setting	12:05 - 12:50
Lunch	12:50 - 1:55
Third Standard-Setting Session	2:00 - 2:45
Stretch-Break	2:45 - 2:50
Instruction in Content and Relevance Judgments	2:50 - 3:00
Content and Relevance Judgments	3:00 - 3:05
Dismissal and Debriefing of Public School Teachers. Break and Snack for University Faculty	3:05 - 3:20
Instructional Validity Instruction and Judgments	3:20 - 3:35
Dismissal & Debriefing of University Faculty	3:35 - 3:50

SCRIPT FOR INTRODUCTORY SESSION
ESSAY TEST

Good morning, I'd like to welcome you to this first session of UNC-Greensboro's Quality Assurance Project.

INTRODUCE YOURSELF, INCLUDING YOUR INSTITUTIONAL AFFILIATION AND YOUR AREA OF SPECIALTY, AND THEN INTRODUCE YOUR ASSISTANT.

Before we get started with some background on today's activities, I'd appreciate it if each of you would take just a moment to introduce yourselves. We'll be doing some work as a group later today, and it will be helpful to know each other.

DESIGNATE A STARTING PERSON AND HAVE THE GROUP INTRODUCE THEMSELVES.

Thank you. We appreciate your willingness to consult with the this North Carolina Quality Assurance Project. The QAP program will substantially change the way we select and certify classroom teachers in North Carolina, and that's likely to have a larger effect on our schools in the next few years than anything else we might imagine. It's important that we contribute to the program and watch its development carefully.

Some of you might not be familiar with the full scope of the Quality Assurance Program, so I'd like to take a few minutes to review it, to place the component we'll be dealing with today in the context of the full program, and to tell you a bit about the National Teacher Examinations.

As you may know, the Quality Assurance Program was first formalized in a joint resolution of the North Carolina State Board of Education and the University of North Carolina's Board of Governors in 1978. Those Boards recognized that,

SHOW TRANSPARENCY No. 1 ON BOARD RESOLUTION

> the state has the responsibility to provide a strong and appropriate instructional program for all its children

and

> The cornerstone of the instructional program delivery system is the competence of the professional personnel charged with responsibility of operating our schools.

In their resolution, the Boards outline a procedure for ensuring the quality of personnel who would, in the future, be fully certified as teachers in North Carolina's schools. Three steps were proposed: First, to administer a series of screening tests to those who applied for admission to the state's teacher education programs. The tests were to measure applicants' competence in social studies, mathematics, science, literature and fine arts, as well as a variety of communication skills. Second, to administer a set of screening tests to graduates of the state's teacher education programs, prior to granting them initial teacher certification. These tests were to measure the candidates' professional knowledge. Third and finally, to develop a program for evaluating and assisting beginning teachers during their first three years of teaching. This period was to culminate in a final screening, to determine which beginning teachers would be granted full, renewable certification.

You can see that the Quality Assurance Program is both ambitious and complex. We hope it will have a positive effect on North Carolina's public education, which will warrant the state's substantial investment. It's important that each of us provide our best professional advice on the program's design and development.

Our focus today will be solely on reviewing the tests to be used in the first phase of the Quality Assurance Program. That is, tests to be used in selecting applicants for North Carolina's teacher education programs. Other projects funded by the State Board of Education address the other components of the QAP.

The State Board of Education has tentatively adopted two tests for screening teacher education applicants. These tests are part of the revised Core Battery of the National Teacher Examinations (known as the NTE), and were developed by Educational Testing Service (called ETS). The Board is awaiting the results of this project before confirming its tentative decision. In particular, we are to tell the Board whether the NTE tests are valid for selecting teacher education students. And if so, we are to define appropriate passing scores that can be used as admission criteria.

Although you will review only one subtest of the NTE Core Battery today, the Writing Essay test, I'd like to take a few moments to tell you about the entire test battery.

DISPLAY TRANSPARENCY SHOWING THE EIGHT TESTS OF THE NTE CORE BATTERY, THE NUMBERS OF ITEMS FOR THE MULTIPLE-CHOICE TESTS AND THE TIME LIMIT FOR THE ESSAY TEST

The State Board of Education has proposed to use two tests from the Core Battery for screening applicants to teacher education programs: the General Knowledge Test and the Communication Skills Test. Each of these tests has four subtests. The General Knowledge Test consists of a 25-item mathematics test, a 30-item science test, a 30-item social studies test, and a 35-item literature and fine arts test. These tests are all in multiple-choice format. ETS computes a single scaled score for the General Knowledge tests, using scores on all four subtests.

The Communication Skills Test consists of a 40-item listening test, a 30-item reading test, and a 45-item writing test, all in multiple-choice format, in addition to a 30 minute writing essay test. As is true for the General Knowledge Test, ETS reports a single scaled score for the Communication Skills Test.

I've aready mentioned that our purpose is to recommend whether these eight subtests are valid for screening teacher education applicants and, if so, what the passing scores on these tests should be. Now, let me explain the process we will use to derive these recommendations.

We will consider three types of test validity. The first two types of validity we'll discuss today might be termed "content validity" and "judgmental predictive validity." We are going to ask each of you whether you judge the NTE Essay Test to be a satisfactory measure of general writing ability. Then we will ask you to judge the relevance of the skills assessed by the writing essay test to success in a teacher education program.

The third type of validity, "instructional validity," will be addressed only by those of you who are faculty members at colleges and universities. Since faculty know the general education curricula at their institutions, we will ask you to tell us whether teacher education applicants would have had the opportunity to learn what the essay test assesses, either as a college prerequisite or during their college studies at your institution.

Although we will restrict ourselves to these three types of validity, let me assure you that we recognize the importance of other forms of test validation. We have recommended to the State Board that this project be the first of several studies of NTE validity. For example, assuming the Board uses the NTE for screening teacher education applicants, it will be important to examine the relationship between applicants' NTE scores and their later performances in the state's teacher education programs.

Now let's talk about standard-setting. In addition to advising us on test validity, we will ask all of you to help formulate a recommended passing score for the writing essay test. During that process, you will review and judge some short essays produced by North Carolina teacher education applicants when they took the NTE Core Battery last spring. I'll give you more specific information about the standard-setting process later on.

DISPLAY TRANSPARENCY SHOWING THE SCHEDULE OF
ACTIVITIES FOR THE DAY

First, let's review the day's schedule.

I'll take about 15 minutes to describe the writing essay test and how it was administered. Then I'll describe the procedures for determining your recommended passing score for the test. That should take no more than 10 minutes.

We estimate that our first standard-setting activity will take about 40 minutes, by which time it will be about 11:30. We'll take a short break. There will be some refreshments in Room __________ .

We'll resume work in this room about 11:40. Prior to the second standard-setting activity, you'll review some actual test performance data for North Carolina teacher education applicants. We'll then have a second standard-setting session that will take about 15 minutes.

The final standard-setting activity, beginning about noon, will be in two sessions. One session will be before lunch and the other after lunch. Each will take about 45 minutes. We'll break for lunch at about 12.50, and return to this room at 2:00. Lunch will be served in __________ . By 2:45 we hope to have all of your recommendations on appropriate standards for the essay test, and we'll take a very short stretch break right in this room.

After our break, I'll tell you in greater detail how we'd like you to report your judgments on the relevance of the skills assessed by the essay test to success in a teacher education program. Then we'll ask you to make those judgments. That will probably take us close to 3:00 o'clock.

Just after three, we'll split the group. The college and university folks will get a 15-minute break -- there will be refreshments in Room ________ . The public school representatives will remain here for their last session. It will consist of some evaluation and housekeeping tasks (such as completing some payment forms for honoraria), and an open discussion of the essay test and the standard-setting procedures.

At 3:20, the college and university faculty will return for their session on instructional validity.

We'll then ask college and university faculty to bear with us through a final session, where we'll ask you for the same kinds of information we sought from public school teachers earlier in the day--- your judgments on the essay test and your reactions to our validation and standard-setting procedures.

We should be finished with everything no later than 4:00 o'clock.

Are there any questions?

ANSWER QUESTIONS AS NEEDED
THEN BEGIN THE TEST FAMILIARIZATION SESSION

Test Familarization Session (10:20 - 10:35)

We've described the Quality Assurance Program, the NTE Battery and today's agenda. Now, let's talk about the Essay Test in some detail. When the two-hour Communication Skills Subtest is administered, the Essay test is given in the last half-hour. That is, students would have already taken three multiple choice tests in Listening, Reading and Writing, each of which take 30 minutes. The multiple choice Writing Test is a 45 item test designed to measure students' ability to use standard written English. Examinees are not tested on grammatical terminology; they are asked to detect errors or choose the best way to rewrite certain phrases or sentences. The Essay test evaluates students' ability with a writing sample .

I'm passing out a copy of the Essay test question and the instructions that are given to the examinees.

ASSISTANT DISTRIBUTES ESSAY QUESTION AND INSTRUCTIONS
SHOW TRANSPARENCY #4 - ETS STATEMENT

While the materials are being distributed, I will read this statement which ETS has requested that we announce:

You are requested not to discuss the tests during data collection.

We have distributed to you current sections of NTE tests - this material is secure.

To maintain the security of the tests

- you are requested not to remove them from the data collection site; if you leave the room please leave all test material at your seat

- you have received a pre-assigned unique serial numbered test packet - this enables us to identify the recipient of any missing test packets.

- you are asked to make no notes or transcripts of the test questions; please do not write on the tests.

- when you are finished, your test packet must be collected. To maintain the necessary security, the materials distributed to you will need to be collected before you leave the data collection room.

ETS is not conducting this validation study. If, however, you have any comments you wish to make regarding the tests, we will be happy to share them with ETS Program and Test Development staffs. We will distribute some sheets for your suggestions.

Now back to our procedures. The directions given to the student are fairly simple; I'll read them.

You will have 30 minutes to plan and write an essay on the topic assigned below. Read the topic carefully. You will probably find it best to spend a little time considering the topic and organizing your thoughts before you begin writing. DO NOT WRITE ON A TOPIC OTHER THAN THE ONE SPECIFIED. AN ESSAY ON A TOPIC OF YOUR OWN CHOICE WILL NOT BE ACCEPTABLE.

An essay question is included in this test to give you an opportunity to demonstrate your ability to write effectively. You should, therefore, take care to express your thoughts on the topic clearly. Be specific, using supporting examples whenever appropriate. Remember that how well you write is much more important than how much you write.

The student writes the essay on a single sheet of paper ruled on two sides and is asked to write legibly. The student is reminded that only one answer sheet will be provided and that in order to fit the whole response on the sheet he/she should not skip lines, or use wide margins.

The essay topics which are selected for the test do not require specialized knowledge by the examinees. The topics draw upon experiences and situations that are common to all members of the general educational community.

Most of the students who take the Communication Skills test will know in advance that an Essay test is required. Of course no student will know the specific topic in advance of the test date. Students may not use dictionaries or other reference tools.

Many students will have read the descriptive brochures on the test which are distributed by ETS. They will know that experienced teachers will grade the papers holistically using procedures which facilitate fair and reliable grading.

Let's describe the holistic scoring procedure. Graders assign a single number on a 1 - 6 scale to represent the overall quality of the essay. A score of "6" represents the highest quality essay and a score of "1" represents the lowest quality essay.

Each essay is scored independently by two teachers experienced in grading student essays. If the grades assigned to the essay are divergent a third rater is employed. ETS then adds the two scores together to produce a score on a new (2 - 12) scale.

The graders are trained to evaluate the essays with a set of "model essays" which are selected from essays written at another test administration. The models serve as guides in defining the points on the scale and represent each possible score point (1 to 6). Prior to the rating session, ETS staff and two composition teachers experienced in essay scoring select the models for the training session. This group must agree on the score given to each model.

Are there any questions at this point?

ANSWER ANY QUESTIONS

If there are no more questions we will begin our first activity in standard setting.

READ THE FOLLOWING:

The purpose of this session is to seek your advice on an appropriate passing score for the Essay portion of the NTE writing test. Let me describe the procedures we'll ask you to follow for determining that score. For now, I'll describe the entire procedure without going into too much detail.

In a moment [Assistant] will pass out the following materials: 1) The question that was used by ETS in testing a large sample of North Carolina's teacher education applicants this past April. 2) a set of eleven essays that were written by that group. Each of the essays has the name of a tree, such as BIRCH, FIR, and so on, printed in the upper right-hand corner, to identify it. 3) A computer-scorable response sheet that has the essay codes printed to the right of several columns of bubbles. The columns are labeled "A" for minimally acceptable, and "U" for unacceptable.

The essays you receive will be stapled together. Because the essays contain confidential test information, we must collect them when you have finished making your judgments. We must also ask that you leave the essays stapled together. We will ask you to read through the entire set of essays in the order they appear in your packet and then, in that order, judge whether or not each essay:

Satisfies your definition of a minimum standard of acceptable writing for a student who is just beginning a teacher education program in North Carolina, When you have completed your judgments for all eleven essays, we will collect that set and your response sheet. We will then give you a second set of eleven essays and a new response sheet. We'll ask you to make the same kinds of judgments for the second set as you did for the first set.

You will be asked to provide initial judgments on the acceptability of the essays based solely on their own merits, and according to your own criteria for clear, effective writing. Of course we want you to keep in mind that the applicants had only 30 minutes to plan and write their essays, that they saw the required essay topic for the first time at the beginning of the 30-minute period, and that they didn't have access to any reference materials, such as a dictionary. Consider these conditions when you judge the essays.

We are asking you to judge two sets of essays rather than one, in order to increase the reliability of our standard-setting procedures. None of the essays is longer than two pages, and most are shorter than that, so reading the essays shouldn't take very long.

Both sets of essays you receive will have been scored by trained raters at Educational Testing Service, and will represent the full range of scores that ETS applies in grading the essay portion of the NTE Writing Test. We want you to consider the essays without being influenced by ETS ratings Once you have provided _your_ judgments, we will show you the scores assigned to the essays by the ETS raters, and we'll discuss the score scale used by ETS. In addition, we will distribute some data on the performance of the North Carolina students who were tested last spring.

At that point we will ask you to reconsider each set of essays. Starting with the first set, we'll ask you to write the ETS-assigned score on each essay. Then we'll ask you to make separate lists of the ETS-assigned scores on the essays you judge to be acceptable and the essays you judge to be unacceptable. After that, we'll ask you to write down the lowest score among the essays you judge to be acceptable, and then the highest score among the essays you judge to be unacceptable. You'll be asked to go through the same process for the second set of essays.

After you have reconsidered the essays and recorded these scores, we will ask you to share with us, and with each other, the factors you considered in judging each essay either "acceptable" or "unacceptable".

Finally, we will ask you to think about the points brought out in the discussion, and look at each set of essays one more time. For each set, you'll then be asked record the lowest score among the essays you judge to be acceptable and the highest score among the essays you judge to be unacceptable, just as you did before.

I'll give more detailed instructions for recording your responses as we go through the various phases of standard-setting. Are there any questions before we begin?

ANSWER QUESTIONS ON PROCEDURES, AS NECESSARY
ANSWER QUESTIONS ON SPECIFIC CRITERIA TO BE APPLIED IN
JUDGING ESSAYS BY REMINDING JUDGES THAT WE WANT THEM TO USE
THEIR OWN CRITERIA, RATHER THAN THOSE USED BY ETS
OR CRITERIA THAT WE INVENT

First Standard-Setting Session (10:45 - 11:25)

READ THE FOLLOWING:

[Assistant] is now distributing the first set of eleven essays and a form for recording your judgments.

DISTRIBUTE ESSAY SET 1, COMPUTER-SCORABLE RESPONSE FORM
AND ESSAY QUESTION

DISPLAY TRANSPARENCY OF RESPONSE FORM -- NO. 5

Please write your name, your social security number, and your school, in the spaces provided for that purpose. Be sure this information IS COMPLETELY OUTSIDE THE AREA ON THE FORM THAT HAS PRINTED BUBBLES. We are asking you to put your name on the form so that we can return it to you during later judgment sessions. Also write your name on your set of essays.

DISPLAY TRANSPARENCY NO. 6

Before you make a judgment about the acceptability or unacceptability of any essay, _read the entire set_ in the order given, so that you can become familiar with the entire range of students' responses. After you have read each essay, then go back to the first one, and, _considering the circumstances under which it was written_, darken the bubble in the "A" column, to the left of the "DOGWOOD" code name, if you want to indicate that the essay:

Is _acceptable_ under _your_ definition of the minimum standard of writing quality you would require of a beginning teacher education student.

Darken the bubble in the "U" column if you want to indicate that the essay:

Is unacceptable under your definition of a minimum standard of writing quality.

You should also mark your decision on each essay. You will be using this information later. Then go on to make the same judgment for each essay in the set. When you finish, please check your response sheet to make sure that you have marked one bubble for each essay. Please raise your hand when you have marked your judgments for all eleven essays. We'll give you the second set of essays and a new response sheet. Please fill in the information requested at the top of the response sheet, and then make the same kinds of judgments for the essays in the second set as you did for those in the first. When you've finished rating all the essays in the second set, let us know and [   ] will collect them. Right after this session we're going to take a break until 11:40, when we'll meet back in this room. So feel free to go to Room ______ for refreshments as soon as you've finished judging the second set of essays and [   ] has collected your materials. Remember we are asking that you NOT discuss the essays or your ratings with anyone during the break. We want to save the discussion for a later session.

Are there any questions before we begin?

ANSWER QUESTIONS AS NECESSARY, AND THEN ASK JUDGES TO BEGIN

AS EACH JUDGE FINISHES, COLLECT HIS/HER ESSAYS AND RESPONSE SHEET, AND GIVE HIM/HER THE SECOND SET OF ESSAYS AND A NEW RESPONSE SHEET.

COLLECT THE SECOND SET OF ESSAYS AND THE RESPONSE SHEET

AS JUDGES SIGNAL YOU THAT THEY HAVE FINISHED

HAVE JUDGES TAKE A BREAK WHEN THEY HAVE FINISHED RATING THE SECOND SET OF ESSAYS

Second Standard-Setting Session (11:50 - 12:05)

DISTRIBUTE ESSAY SET NO. 1, ESSAY RATING FORM FOR SECOND JUDGMENT SESSION (For Essay Set No. 1), AND GRAPH FOR THE ESSAY TEST

READ THE FOLLOWING:

We're now ready to begin our second essay rating session. [ ] is redistributing the first set of essays you judged, an ESSAY RATING FORM that includes the scores the ETS readers gave those essays, and a graph that shows the distribution of performance of teacher education applicants in North Carolina, when they took the Essay test last April. We'll look at these together in a few minutes. Please write your name, your social security number, and your school, college or university in the spaces provided at the top of the ESSAY RATING FORM.

DISPLAY TRANSPARENCY OF ESSAY RATING FORM -- No. 7

GIVE PARTICIPANTS A MINUTE TO COMPLETE THE TOP OF THE FORM

Before you begin the next rating task, we'll discuss some features of the ETS rating scale and the information in the graph. ETS had two independent readers score each essay on a scale of 1 to 6, with "6" being best and "1" being poorest. A final score for each essay was determined by summing the scores assigned to the essay by these readers.

Thus if both readers gave an essay a score of 1, it received a final score of 2. If one reader gave an essay a score of 2 and the other gave it a score of 3, it received a final score of 5. If both readers gave an essay a score of 6, it received the highest possible final score of 12. If the two readers disagreed by more than one point in the scores they applied to an essay, a third reader was used to resolve the disagreement. None of the essays you have seen today fell into this category. Each of the two sets of eleven essays contains one essay at each end of the eleven scale points from 2 through 12.

Are there any questions?

ANSWER QUESTIONS ON THE ETS SCORING PROCEDURE AS NECESSARY

DEFER QUESTIONS ON THE CRITERIA USED BY ETS RATERS
IF NECESSARY, TELL JUDGES THAT WE WANT THEM TO USE
THEIR OWN CRITERIA

When there are no more questions

DISPLAY TRANSPARENCY OF GRAPH -- No. 8

Now let's consider the information in the graph you received a few minutes ago. We constructed the graph from data on the test performance of nearly 700 North Carolina teacher education applicants. The data were collected in a special administration of the NTE Revised Core Battery this past spring.

The horizontal scale of the graph shows all of the possible values on the ETS score scale (2 to 12). The vertical scale of the graph shows percents, going from zero to 100. The graph has been plotted to show the percentage of

North Carolina applicants who would have failed the test last spring, had the minimum passing score been set at each of the scale points from 2 to 12. For example, had the minimum passing score been set at 4, 7 percent of students would have failed the Essay test and 93 percent would have passed. You can verify the percent who would have failed by finding 4 on the horizontal scale of the graph, drawing a vertical line up to the curve, and then drawing a horizontal line over to the percent scale from the point where your vertical line crosses the curve.

DEMONSTRATE THE METHOD FOR FINDING THE PERCENT WHO WOULD HAVE FAILED WITH A STANDARD OF 4 BY USING TRANSPARENCY No. 8

You can find that 93 percent would have passed by subtracting 7 from 100 percent.

Try to find the percent who would have failed the test and the percent who would have passed the test had the standard been set at 8. What are the percentages?

ALLOW TIME FOR JUDGES TO FIND ANSWERS

Right, the percent who would have failed is 48 and the percent who would have passed is 52. Did anyone get different answers? Are there any questions on using the graph?

ANSWER QUESTIONS AS NECESSARY

At this point, you have a table that shows the scores ETS readers assigned to each of the essays in the first set.

In addition, you can use the graph to determine the percentage of North Carolina applicants who would have failed the Essay test, had the passing score been set at any value between 2 and 12. Assume the standard is to be set at the lowest score of the essay you judged to be acceptable. We would like you to think about this information and reconsider your earlier judgments on the acceptability of each essay. You'll then use a new method to recommend a minimum passing score for the Essay test. Let me describe the process we'd like you to follow:

DISPLAY TRANSPARENCY OF ESSAY RATING FORM -- No. 7

You will find a table on the Essay Rating Form that lists the final score assigned to each of the essays in Set No. 1 by the ETS readers. Please take a moment to write the appropriate ETS score from that table in the upper right hand corner of each essay. Put a 2 on the DOGWOOD essay, a 3 on the OAK essay, and so on.

GIVE PARTICIPANTS TIME TO WRITE THE SCORES ON THE ESSAYS

Now please read through the essays once again to refresh your memory. After you have re-read the essays, please reconsider your earlier judgments on whether each essay is or is not acceptable under your definition of the minimum standard of writing quality. Once again, keep in mind the conditions under which the essays were written. You might also want to consider the information in the graph and the scores assigned to the essays by the ETS readers. However, do not feel that you have to change any of your earlier judgments.

Do not be concerned that ETS readers might have assigned a relatively high score to an essay that you judged to be unacceptable. If this happened, it probably means that ETS readers used criteria for judging the quality of the essays that were somewhat different from yours. Since we are interested in _your_ judgments, this is not a problem. We want you to apply your own criteria. If we gave you the ETS criteria, that might influence your judgement.

Are there any questions?

ANSWER QUESTIONS AS NECESSARY

As you reconsider your judgements, note any changes on the essays. When you've gone back over Set No. 1, please record your judgements on the Essay Rating Form in this way:

Write the ETS scores _only_ of the _essays you judged to be acceptable_ in the spaces right below the table of ETS scores. Don't write the scores of _unacceptable_ essays in these spaces.

When you have finished, please find the _lowest_ of the ETS scores you marked in the spaces. On the row of numbers below the spaces, circle the number that corresponds to the _lowest_ of the ETS scores on the "acceptable" essays.

Next, please look back through the essays to find the ETS scores of the _essays you judged to be unacceptable_. Right below the next statement on the form, you will find another set of spaces. Please use those spaces to write the ETS scores of the essays you judged to be unacceptable.

Finally, please use the bottom row of numbers from 2 through 12 to indicate the highest ETS score assigned to an essay you judged to be unacceptable. Just circle the corresponding number in the row at the bottom of the form.

Are there any questions?

ANSWER QUESTIONS AS NECESSARY

When you have finished your review of the first set of essays and you've completed your response form, let us know. We will then exchange the first set of essays and response forms for your second set and a new response form. Handle these in the same way. Fill in the information requested at the top of the ESSAY RESPONSE FORM. Then use the table of scores given by ETS readers and write the appropriate ETS score on each essay. From that point on, please follow the same judgment procedures as you reconsider your earlier decisions on the acceptability of each essay. Record your judgments on the new ESSAY RATING FORM in the same way you completed the first one.

When you've finished with the second set of essays, please raise your hand and we'll collect your materials. You may relax at your desk or leave the room as you wish. We'll be starting our last session before lunch as soon as everyone has finished with the second set of essays. Don't wander out of earshot.

Are there any questions?

ANSWER QUESTIONS AS NECESSARY

Please begin.

AFTER YOU'VE COLLECTED ALL OF THE ESSAYS AND BOTH RESPONSE SHEETS FROM ALL JUDGES, PLEASE START THE LAST STANDARD-SETTING SESSION BEFORE LUNCH

Third Standard-Setting Session -- Part 1 (12:05 - 12:50)

DISTRIBUTE ESSAY SET No. 1
AND A NEW ESSAY RATING FORM (Third Session - Set No. 1)

We're now ready to begin our last session before lunch. [ ] is handing out the first set of essays one last time, together with a new ESSAY RATING FORM. Please raise your hand when [Assistant] calls your name. Take a moment to fill in the information requested at the top of the new ESSAY RESPONSE FORM.

At this point, I would like you to look at the first essay in the set, the one marked DOGWOOD.

GROUP LEADER NOTE:

YOU WILL BE GIVEN A PRINTOUT THAT SHOWS THE NUMBER OF RATERS WHO MARKED "A" AND "U" ON EACH ESSAY
ADAPT THE SCRIPT AS NECESSARY, TO FIT THESE DATA.

[Several of you judged this essay to be acceptable and several of you rated it unacceptable] [None of you judged this essay to be acceptable]. [Would someone who judged this essay acceptable be willing to discuss your judgment -- share with the group, some of the reasons you felt the essay was acceptable?] [Feel free to make notes on your essay copy.] [Now would someone who judged the essay unacceptable tell us why you felt it didn't measure up?] Again, take notes if you'd like.

USE ALL OF YOUR LEADERSHIP SKILLS AND CHARM HERE. ENCOURAGE DISCUSSION. PROBE AND CLARIFY AS NECESSARY. MAKE SURE THAT YOU DON'T INJECT YOUR REASONS, REGARDLESS OF THE TEMPTATION. WHEN THEY'VE FINISHED WITH THE FIRST ESSAY GO ON TO DISCUSS THE REST, IN ORDER. KEEP ENCOURAGING PEOPLE TO MAKE NOTES. YOU WILL BE GIVEN A TABULATION OF THE NUMBER OF PEOPLE WHO MARKED EACH ESSAY "A" AND THE NUMBER WHO MARKED EACH ESSAY "U", SO YOU'LL KNOW WHETHER YOU HAVE BOTH JUDGMENTS FOR EACH ESSAY. DON'T TRY TO PRODUCE AGREEMENT. JUST GIVE EACH "SIDE" AN ADEQUATE "HEARING."

READ THE FOLLOWING:

We've now discussed all eleven essays in the first set. Although each of us might have different reasons for judging an essay to be acceptable or unacceptable, we how have a better idea of the criteria applied by others in the group. In light of this new information, please reconsider your initial judgments on each essay. Then use the ESSAY RESPONSE FORM to record your judgments. Please use the same procedures you used in our second rating session to complete the form. That is, write the ETS scores of the essays you judge to be acceptable in the appropriate spaces. Then circle the number of the acceptable essay with the lowest ETS-assigned score. Next, write the ETS scores of the essays you judge to be unacceptable in the bottom row of spaces, and then circle the number in the bottom row for the unacceptable essay with the highest ETS-assigned score. When you've finished we'll collect your essays and your response forms. When you have given me your essays and your response forms, you might want to go right on to lunch. Lunch is being served in Room ______. We'll meet back in this room at 2:00 o'clock for a final standard-setting session using the second set of essays. Again I must remind you to not discuss the essays during lunch.

MAKE SURE YOU COLLECT ALL OF THE ESSAYS AND RATING FORMS BEFORE PEOPLE LEAVE THE ROOM

LUNCH (12:50 - 1:55)

Third Standard-Setting Session - Part 2 (2:00 - 2:45)

DISTRIBUTE ESSAY SET No. 2, ESSAY RATING FORMS FROM SESSION 1 (SET No. 2), and NEW ESSAY RATING FORM (FOR ESSAY SET No. 2 -- THIRD SESSION)

READ THE FOLLOWING:

I will now ask you to repeat the final judgment activity for the second set of essays. [   ] is handing out the second set of essays, the judgment sheets you completed during the first standard-setting session, and the last ESSAY RATING FORM we'll ask you to complete today. Please raise your hand when [Assistant] calls your name and institution. Please fill in the information at the top of the ESSAY RESPONSE FORM. Then take a moment to review your earlier judgments on the second set of essays.

ALLOW TIME FOR JUDGES TO LOOK OVER THE ESSAYS

PROCEED AS BEFORE, ADAPTING SCRIPT AS NECESSARY TO FIT THE DATA YOU HAVE BEEN GIVEN

At this point, I would like you to look at the first essay in the set, the one marked COFFEE. [Several of you judged this essay to be acceptable and several of you rated it unacceptable] [None of you judged this essay to be acceptable]. [Would someone who judged this essay acceptable be willing to discuss your judgment -- share with the group, some of the reasons you felt the essay was acceptable?] [Feel free to make notes on your essay copy.] [Now would someone who judged the essay unacceptable tell us why you felt it didn't measure up?] Again, take notes if you'd like.

USE ALL OF YOUR LEADERSHIP SKILLS AND CHARM HERE. ENCOURAGE DISCUSSION. PROBE AND CLARIFY AS NECESSARY. MAKE SURE THAT YOU DON'T INJECT YOUR REASONS, REGARDLESS OF THE TEMPTATION. WHEN THEY'VE FINISHED WITH THE FIRST ESSAY GO ON TO DISCUSS THE REST, IN ORDER. KEEP ENCOURAGING PEOPLE TO MAKE NOTES. YOU WILL BE GIVEN A TABULATION OF THE NUMBER OF PEOPLE WHO MARKED EACH ESSAY "A" AND THE NUMBER WHO MARKED EACH ESSAY "U", SO YOU'LL KNOW WHETHER YOU HAVE BOTH JUDGMENTS FOR EACH ESSAY. DON'T TRY TO PRODUCE AGREEMENT. JUST GIVE EACH "SIDE" AN ADEQUATE "HEARING."

READ THE FOLLOWING:

We've now discussed all eleven essays in the second set. In light of this new information, please reconsider your initial judgments on each essay, and once again, use the ESSAY RESPONSE FORM to record your judgments. Please use the same procedures you used in our session right before lunch to complete the form. That is, write the ETS scores of the essays you judge to be acceptable in the appropriate spaces. Then circle the number of the acceptable essay with the lowest ETS-assigned score. Next, write the ETS scores of the essays you judge to be unacceptable in the bottom row of spaces, and then circle the number in the bottom row for the unacceptable essay with the highest ETS-assigned score. Please raise your hand when you've finished, and we'll collect your essays and your response forms.

WHEN FORMS HAVE BEEN COLLECTED FROM EVERYONE, READ THIS:

This ends our standard-setting activities for the Essay test. We appreciate your thoughtful contributions today. Recommending test standards is a difficult task, and you've given us information that will help tremendously when we formulate our recommendations to the State Board of Education.

Let's take a short break right here in the room before we turn to our first validation task.

Instruction in Content and Relevance Judgments (2:50 - 3:00)

A fundamental question that must be answered before a test is used for screening is: Is the test valid for the purpose for which it is to be used?". The NTE Essay Test has been tentatively designated to be used as part of the QAP screening procedures because it is assumed that the writing skills assessed by the this essay test are important to success in teacher education programs. This version of the NTE has not been validated for screening applicants to teacher education programs. Therefore two initial questions must be answered.

First, we must ask whether this NTE Essay examination is a satisfactory measure of writing ability. At this point in the session you have a pretty complete picture of the NTE Essay test. We would like your judgment of whether this test measures general writing ability. We will call this type of validity "content validity."

Second, we must ask if the skills measured by this essay test are important to success in a teacher education program. That is, are the writing skills assessed by this test relevant to successful performance in teacher training? This type of validity we will call "judgmental predictive validity."

It is important to recognize that the behavior facilitated by the possession of writing skills is success as a student in a teacher education program and not success as a teacher in the classroom. We agree that writing skills probably are important to teacher success in the classroom, but that is not the focus of the current study. Another study has been proposed which will examine the relationship between NTE Test scores obtained when a student applies for admission to teacher education and subsequent measurements of the student's academic performance. In the current study we are seeking the judgment of experts concerning whether the NTE Essay Test actually measures writing ability and whether writing ability is important to success in teacher education.

[NAME] is distributing a two page response form which you should use to record your validity judgments.

DISTRIBUTE ESSAY TEST CONTENT /RELEVANCE JUDGMENT RESPONSE FORM
(form1 and form2)

First, please write your name and institution or school name in the spaces provided on the response sheet.

Having already reviewed, in detail, the Essay Test administration procedures and scoring procedures used by Educational Testing Service, let's read through the two main questions on the response form.

SHOW TRANSPARENCY OF ESSAY CONTENT /RELEVANCE JUDGMENT RESPONSE FORM #1

Question #1

Please read the following statement and indicate whether you Strongly Agree(SA), Agree(A), are Undecided (U), Disagree(D), or Strongly Disagree(SD).

"The NTE Essay Examination which I have reviewed, including its administration and scoring procedures, is a satisfactory measure of general writing skill"

Please circle your response:

SA A U D SD

Please describe the reasons for your answer in the space below the question and on the back of the page.

Now let's consider the second question. Assuming that the NTE Essay test does measure writing ability, next we would like to obtain your judgment on the importance of the writing skills it measures to success as a student in a teacher education program. The second question asks:

Question #2

"Please consider the writing skills measured by the NTE Essay Test and provide us with your judgment of how important those skills are to success in the teacher education program with which you are most familiar. Specifically, we would like to know how important it would be to have minimal writing skills as measured by the NTE Essay Test.

Please respond by placing an 'X' on the line in front of one of the following choices. Your response sheet contains these definitions:

SHOW TRANSPARENCY OF ESSAY CONTENT/RELEVANCE
JUDGMENT RESPONSE FORM (form 2)

"Crucial" -of critical importance to success in a teacher education program. These are skills or knowledge which a student would certainly need in order to be successful in the teacher education curriculum with which you are most familiar.

"Important"- of great value to success in a teacher education program. These include skills or knowledge which if present in a student would certainly contribute to a successful experience in teacher education; however it would be possible for a student without some of these skills and knowledge to succeeed in teacher education.

"Of Questionable Importance"- these include skills or knowledge which might enhance performance in teacher education but which are not essential to success.

"Irrelevant to Success in Teacher Education"- have no value or effect; these skills or knowledge have no bearing whatsoever on successful performance in teacher education.

You'll notice that there is an open-ended question following the importance rating. The question says:

"In the space below please elaborate on your response to the question which you answered above. That is, describe the reasons you used to make your choice.

Content and Relevance Judgments (3:00 - 3:05)

It is now 3:00 pm. Please begin to answer the questions. I think that most people should finish by 3:05. If you finish early, please raise your hand and we'll pick up your materials.

AT 3:05 SAY

Is everyone finished? [If not please begin to finish] [NAME] will pick up your materials.

ASSISTANT PICKS UP ALL MATERIALS

At this point the college and university faculty will take a fifteen minute coffee or coke break. Please return to this room by 3:20.

WAIT WHILE COLLEGE FACULTY LEAVE THE ROOM

The public school faculty are finished with all the major judgments we're going to ask of you. If you'll remain in the room for a few more minutes we would like to get your reactions to the process we've just been through and to the NTE. We'll answer any questions you have.

DISTRIBUTE EVALUATION QUESTIONNAIRE,COMPUTER READABLE RESPONSE FORM, INFORMATION AND PAYMENT FORMS

Our last activity will take only a short time. [NAME] is distributing an Evaluation Form. There is also an Information Form and a Payment Form to be completed.

Please take a few moments to complete the Information form and payment forms so we can send you your honorarium. We must ask you to complete the top portion of two copies of a blue payment form, and that you sign each form.

Please use the green, computer-readable PARTICIPANT'S EVALUATION RESPONSE FORM to provide your responses to the questions on the PARTICIPANT'S EVALUATION QUESTIONNAIRE. Please be sure to write your name and your social security number at the top of the RESPONSE FORM. Here again, we will respect the confidentiality of your judgments, and use your responses only for data analysis purposes. We will not identify your individual responses in any of our reports.

ALLOW TIME FOR COMPLETING FORMS

WHEN IT APPEARS THAT EVERYONE HAS FINISHED,

READ THE FOLLOWING:

Has everyone completed an evaluation form, an INFORMATION FORM, PARTICIPANT'S EVALUATION RESPONSE FORM? [ ] will collect your forms.

This completes our formal standard-setting and validation activities. Do any of you have any comments on either the NTE [ ] test or on the procedures we've used today? I've asked [ ] to record your comments so that we can either pass them on to ETS or consider them ourselves when we redesign our procedures.

ALLOW TIME FOR DISCUSSION

Instructional Validity (3:20 - 3:35)

There is one more task which only you as college and university faculty can perform. It relates to the issue of validity which we discussed earlier.

One concern that has been raised in the courts with respect to the use of screening procedures has come to be called instructional validity. The courts have held that when screening procedures are used, they must be "fair". If the Quality Assurance Program policy is implemented as it is currently planned, it says that students will have to demonstrate the ability to write an essay at a satisfactory level of quality before they may be admitted to teacher education programs. For that to be a fair requirement, we must be sure that applicants have had a reasonable educational opportunity to master the skills necessary to write a minimally acceptable essay. For that reason, we will ask you to think about the skills necessary to produce a minimally acceptable essay for the NTE and then consider whether most of the students at your institution, when they apply for admission to teacher education, would have had a reasonable opportunity to acquire those skills.

ASSISTANT DISTRIBUTES ESSAY INSTRUCTIONS, ETS SPECIFICATIONS FOR WRITING TEST, INSTRUCTIONAL VALIDITY RESPONSE FORM(1 and 2)

We are distributing a copy of the Essay Test Instructions, a sheet listing the ETS Specifications for the Writing test and an Essay Test Instructional Validity Response Form.

First let's take a look at what Educational Testing Service describes as effective writing. This statement summarizes the view of the ETS Writing Subcommittee on what the effective writer should be able to do and what the teaching of effective writing should emphasize. This statement was used as a guide in constructing both the multiple choice and essay writing tests. The abilitites are:

1. provide and sustain a focus or thesis
2. attain in different papers the varied aims or purposes (e.g., explanatory, persuasive, expressive) of discourse
3. decide which of these aims or purposes is appropriate in a given writing situation
4. select and sustain an appropriate persona or voice
5. produce and develop adequate and appropriate materials to accomplish the purpose for writing, identifying and supporting, as appropriate, important assumptions
6. choose and use a mode of organization consistently
7. preserve coherence in an extended piece of writing
8. choose an appropriate mode of organization (chronological, enumerative, etc.)
9. construct sentences in standard written English, adjusting choice of sentence structure and word choice to suit purposes and aims
10. use sentences and vocabulary which are appropriate to the purpose of the writing
11. use words and sentences which are appropriate for the intended readers
12. construct sentences in standard written English and identify sentences that do not observe the conventions of standard written English, such as grammar, usage, and punctuation

A few of these skills may not be assessable by the NTE Essay Test. For example, with respect to point # 2, different papers can't be written for the single NTE Essay question. However, the statement taken as a whole does describe the objectives that general instruction in writing skills should have, and it does provide a focus for us to ask the question: " Will students have had a reasonable opportunity to learn the skills described in the ETS statement?"

Please take the sheet marked "Essay Test Instructional Validity Response Form", and write-in your name and institution at the top.

DISPLAY TRANSPARENCY OF INSTRUCTIONAL VALIDITY FORM #1

There are two questions. The statement for the first one reads:

"At least 90% of the students applying for admission to teacher education at the institution with which I am most familiar have had a reasonable opportunity to receive instruction in the Writing skills described in the ETS statement before applying for admission to teacher education."

You are being asked to read this statement and to indicate whether you Strongly Agree(SA), Agree(A), are Uncertain(U), Disagree(D), or Strongly Disagree(SD).

Several points should be made about that statement. We have used the phrase "opportunity to receive instruction." By that, we mean students should have been able to study or take courses in which writing skills would have been taught. We don't mean that all students would necessarily have learned well, but simply that they would have had the opportunity to learn.

Second, when we say the opportunity existed before applying for admission to teacher education we don't necessarily restrict the opportunity for learning to college level studies. The opportunity to learn might have come at some point before the general educational studies of college, for example in high school.

We are asking you to respond by circling the choice which best represents your judgment. After you do that, please write a brief statement in the space provided and describe the reasons for your judgment.

We know this is a difficult global judgment. However please realize that this is the only practical way we have of obtaining this kind of information. Research has shown that faculty judgment is quite accurate. So please do your best to respond to this question.

Please begin to answer this question.

IN ABOUT FIVE MINUTES SAY

Is everyone finished? [If not please begin to finish up.]

WHEN EVERYONE IS FINISHED

In the first Instructional Validity judgment question we asked you to look at the list of specifications that was written for the Essay Test and decide whether students at your institution would have the opportunity to acquire those listed skills through instruction.

It is possible that the Essay test as it actually is administered requires skills that are not those skills described in the ETS statement. Therefore, now we would like you to focus on the manner in which minimal competence in writing is actually assessed by this test. At this point, you know the test procedures and the instructions which are given to students. You also know about the scoring procedures. In fact you know from first-hand experience what scores are applied to an actual set of essays. We want you to use that information to decide which Writing skills are actually assessed by the NTE Essay test. Then we would like you to decide whether students at your institution would have had an opportunity to learn those skills.

On the Essay Test Instructional Validity Response Form there is a second question. It reads:

SHOW TRANSPARENCY OF ESSAY TEST INSTRUCTIONAL VALIDITY
RESPONSE FORM #2

Question 2

Please read the following statement and indicate whether you Strongly Agree(SA), Agree(A), are Undecided(U), Disagree(D), or Strongly Disagree(SD),

"At least 90% of the students applying for admission to teacher education at the institution with which I am most familiar have had a reasonable opportunity to receive instruction in the writing skills that are actually measured by the NTE Essay test."

PLEASE CIRCLE YOUR RESPONSE

SA A U D SD

Please begin by thinking about the skills you believe are actually measured by the NTE, then read the statement and mark the extent to which you agree or disagree. There is a space below the question for you to describe the reasons for your judgment.

WAIT ABOUT FIVE MINUTES FOR JUDGES TO MARK THEIR RESPONSE

WHEN EVERYONE IS DONE SAY

Is everyone finished? [If not please begin to finish up.]

ASSISTANT COLLECTS ESSAY INSTRUCTIONS, ETS STATEMENT, INSTRUCTIONAL VALIDITY RESPONSE FORMS

Debriefing Session and Evaluation (3:35 - 3:50)

DISTRIBUTE EVALUATION QUESTIONNAIRE, INFORMATION AND PAYMENT FORMS

Our last activity will take only a short time. [NAME] is distributing an "Evaluation Form". There is also an Information Form and a Payment Form to be completed.

Please take a few moments to complete the information form and payment forms so we can send you your honoraria. We must ask those of you affiliated with non-public institutions to complete the top portion of two copies of a blue payment form, and that you sign each form. We will be able to secure payment information from the Information Form for those of you who are affiliated with the University of North Carolina.

Please use the green, computer-readable PARTICIPANT'S EVALUATION RESPONSE FORM to provide your responses to the questions on the PARTICIPANT'S EVALUATION QUESTIONNAIRE. Please be sure to write your name and your social security number at the top of the RESPONSE FORM. Here again, we will respect the confidentiality of your judgments, and use your responses only for data analysis purposes. We will not identify your individual responses in any of our reports.

ALLOW TIME FOR COMPLETING FORMS

WHEN IT APPEARS THAT EVERYONE HAS FINISHED,

READ THE FOLLOWING:

Has everyone completed an evaluation form, an INFORMATION FORM, PARTICIPANT'S EVALUATION RESPONSE FORM? Those of you at a non-UNC campus should also make sure you have completed and signed both copies of the blue payment form. [ ] will collect your forms.

This completes our formal standard-setting and validation activities. Do any of you have any comments on either the NTE [ ] test or on the procedures we've used today? I've asked [ ] to record your comments so that we can either pass them on to ETS or consider them ourselves when we redesign our procedures.

ALLOW TIME FOR DISCUSSION

I want to thank each of you for participating in this Quality Assurance Project. Without your help, we couldn't give the State Board sound advice on use of the NTE in North Carolina. It has been a pleasure working with you. Drive carefully on your way home.

MATERIALS AND INSTRUMENTATION FOR MULTIPLE-CHOICE AND ESSAY TESTS

ROSTER FOR TEST BOOKLETS

TEST NAME____________________

		BOOKLET SERIAL NUMBER		
Judge Name	SS Number	Test I	Test II	Test III

PANELIST'S COMMENT SHEET

Name ______________________ Date ______________________

Institution ____

Test ______________________ Panel ______________________

State ____

tem #	Remarks

NORTH CAROLINA TEACHER QUALITY ASSURANCE PROGRAM

ITEM RELEVANCE/ INSTRUCTIONAL VALIDITY JUDGMENT FORM

① PLEASE PRINT THE REQUESTED INFORMATION IN THE APPROPRIATE BLANKS BELOW:

YOUR NAME ____________ EMPLOYED AT (COLLEGE or SCHOOL) ____________ DATE ____________

COLLEGE DEPARTMENT ____________ (OR) SUBJECT/LEVEL TAUGHT ____________ TEST REVIEWED ____________

IMPORTANT:

USE NO. 2 PENCIL ONLY

- EXAMPLE 1 ▬ 3 4 5
- ERASE COMPLETELY TO CHANGE
- MAKE YOUR MARKS HEAVY & DARK

② PLEASE WRITE YOUR SOCIAL SECURITY NUMBER IN THE BOXES BELOW THEN FILL-IN THE BUBBLES BELOW THE LINE FOR EACH DIGIT.

③ IDENTIFY THE NUMBER OF THE TEST YOU ARE REVIEWING IN THE LIST BELOW, THEN FILL-IN THE BUBBLE ON THE LINE INDICATED.

TEST NAME	NUMBER
READING	1
WRITING	2
LISTENING	3
MATHEMATICS	4
SCIENCE	5
SOCIAL STUDIES	6
LIT./FINE ARTS	7

④ MAKE YOUR RELEVANCE JUDGMENTS FOR EACH ITEM ON THE FOLLOWING SCALE, AND FILL-IN BUBBLES ON THE LINES INDICATED BELOW.

4 = CRUCIAL TO SUCCESS IN TEACHER ED.
3 = IMPORTANT TO SUCCESS IN TEACHER ED.
2 = QUESTIONABLE RELEVANCE TO SUCCESS IN TEACHER ED.
1 = NOT RELEVANT TO SUCCESS IN TEACHER ED.
? = I CAN'T MAKE THE JUDGMENT

⑤ MAKE YOUR INSTRUCTIONAL VALIDITY JUDGMENT FOR EACH ITEM ON THE FOLLOWING SCALE, AND FILL-IN BUBBLES ON THE LINES INDICATED BELOW. FILL-IN ONE BUBBLE ON EACH LINE.

Y = YES
N = NO
? = I CAN'T MAKE THE JUDGMENT

② FILL-IN APPROPRIATE BUBBLES TO RECORD SOCIAL SECURITY NO. ON LINES BELOW. (MARK ONE BUBBLE ON EACH LINE)

0	1	2	3	4	5	6	7	8	9
0	1	2	3	4	5	6	7	8	9
0	1	2	3	4	5	6	7	8	9
0	1	2	3	4	5	6	7	8	9
0	1	2	3	4	5	6	7	8	9
0	1	2	3	4	5	6	7	8	9
0	1	2	3	4	5	6	7	8	9
0	1	2	3	4	5	6	7	8	9
0	1	2	3	4	5	6	7	8	9

③ RECORD TEST NO. ON LINE BELOW

1	2	3	4	5	6	7

④ RECORD RELEVANCE JUDGMENTS BELOW

ITEM NO	1	2	3	4	?
1	1	2	3	4	?
2	1	2	3	4	?
3	1	2	3	4	?
4	1	2	3	4	?
5	1	2	3	4	?
6	1	2	3	4	?
7	1	2	3	4	?
8	1	2	3	4	?
9	1	2	3	4	?
10	1	2	3	4	?
11	1	2	3	4	?
12	1	2	3	4	?
13	1	2	3	4	?
14	1	2	3	4	?
15	1	2	3	4	?
16	1	2	3	4	?
17	1	2	3	4	?
18	1	2	3	4	?
19	1	2	3	4	?
20	1	2	3	4	?
21	1	2	3	4	?
22	1	2	3	4	?
23	1	2	3	4	?
24	1	2	3	4	?
25	1	2	3	4	?
26	1	2	3	4	?
27	1	2	3	4	?
28	1	2	3	4	?
29	1	2	3	4	?
30	1	2	3	4	?
31	1	2	3	4	?
32	1	2	3	4	?
33	1	2	3	4	?
34	1	2	3	4	?
35	1	2	3	4	?
36	1	2	3	4	?
37	1	2	3	4	?
38	1	2	3	4	?
39	1	2	3	4	?
40	1	2	3	4	?
41	1	2	3	4	?
42	1	2	3	4	?
43	1	2	3	4	?
44	1	2	3	4	?
45	1	2	3	4	?

⑤ RECORD INSTRUCTIONAL VALIDITY JUDGMENTS BELOW

Y	N	?	ITEM NO
Y	N	?	1
Y	N	?	2
Y	N	?	3
Y	N	?	4
Y	N	?	5
Y	N	?	6
Y	N	?	7
Y	N	?	8
Y	N	?	9
Y	N	?	10
Y	N	?	11
Y	N	?	12
Y	N	?	13
Y	N	?	14
Y	N	?	15
Y	N	?	16
Y	N	?	17
Y	N	?	18
Y	N	?	19
Y	N	?	20
Y	N	?	21
Y	N	?	22
Y	N	?	23
Y	N	?	24
Y	N	?	25
Y	N	?	26
Y	N	?	27
Y	N	?	28
Y	N	?	29
Y	N	?	30
Y	N	?	31
Y	N	?	32
Y	N	?	33
Y	N	?	34
Y	N	?	35
Y	N	?	36
Y	N	?	37
Y	N	?	38
Y	N	?	39
Y	N	?	40
Y	N	?	41
Y	N	?	42
Y	N	?	43
Y	N	?	44
Y	N	?	45

SCANTRON FORM NO. 20[illegible]-UNCG FEED THIS DIRECTION

DIRECTIONS FOR STANDARD-SETTING

In this session we will ask you to make a judgment about item in the [ ] section of the NTE. The type of gment is as follows:

You are to imagine a hypothetical applicant who has the um amount of knowledge and skill in the subject-area of the] test, necessary to succeed in the North Carolina cher education program with which you are most familiar. You t think of this hypothetical applicant as a "low C" student, east in the areas of skill and knowledge assessed by the] section of the NTE.

You are to estimate, for each item, the probability or nces" that this minimally competent applicant would be able answer the test item correctly.

Probabilities are expressed as decimal numbers and represent chances or likelihood that something will occur. A probabil-y can range from .00, meaning "not possible", to 1.00, meaning ertain." A probability of .50 corresponds to a 50% likelihood. you decided the probability that an applicant could answer a st item correctly was .50, you would be saying that he or she d a 50-50 chance of answering the test item correctly.

Since the NTE [ ] test is scored by counting the number of rect answers and has no penalties for wrong answers, an examnee could get full credit on a test item either by knowing the rect answer or by guessing successfully. So when you estimate babilities, remember that an applicant could get the right nswer either through knowledge or with a lucky guess. An examnee with no knowledge of the correct answer who guessed at ndom on a five-option multiple-choice item would have a probaility equal to 0.20 of getting that item right. On a four-option item, the probability of getting the item right by guessing ndomly would be 0.25. Measurement research has shown that some aminees, particularly those who do poorly on multiple-choice tests, do not always guess the answer to an item they do not ow, even when it is to their advantage to do so. Therefore, tual probabilities of success due to guessing are likely to be somewhat lower than the values we have just discussed.

For each item on the [ ] test, record your estimate of he probability of a correct answer, on the JUDGMENT FORM FOR STANDARD-SETTING. Completely darken the rectangular bubble that corresponds to the interval containing your estimate. For xample, if you believe the probability for an item is 0.10 or ess, darken the bubble numbered "1" for that item. The "1" bubble represents all probability estimates from zero through .10. The bubble numbered "2" represents all probability estiates from 0.11 through 0.20. And so on. The number in each bubble corresponds to the probability value at the upper end of the range it represents. If, for example, you estimated the probability value for an item to be 0.56, you would darken the bubble numbered "6" for that item, since your estimate of 0.56 would be between the probability values 0.51 and 0.60 represented by bubble number "6."

NORTH CAROLINA TEACHER QUALITY ASSURANCE PROGRAM

JUDGMENT FORM FOR STANDARD SETTING

① PLEASE PRINT THE REQUESTED INFORMATION IN THE APPROPRIATE

NORTH CAROLINA TEACHER QUALITY ASSURANCE PROGRAM

JUDGMENT FORM FOR STANDARD SETTING

① PLEASE PRINT THE REQUESTED INFORMATION IN THE APPROPRIATE BLANKS BELOW:

YOUR NAME __________ EMPLOYED AT (COLLEGE or SCHOOL) __________ DATE __________

COLLEGE DEPARTMENT __________ (OR) SUBJECT/LEVEL TAUGHT __________ TEST REVIEWED __________

IMPORTANT:

USE NO. 2 PENCIL ONLY

- EXAMPLE ⊂1⊃ ● ⊂3⊃ ⊂4⊃ ⊂5⊃
- ERASE COMPLETELY TO CHANGE
- MAKE YOUR MARKS HEAVY & DARK

② PLEASE WRITE YOUR SOCIAL SECURITY NUMBER IN THE BOXES BELOW THEN FILL-IN THE BUBBLES BELOW THE LINE FOR EACH DIGIT.

③ IDENTIFY THE NUMBER OF THE TEST YOU ARE REVIEWING IN THE LIST BELOW, THEN FILL-IN THE BUBBLE ON THE LINE INDICATED.

TEST NAME	NUMBER
READING	1
WRITING	2
LISTENING	3
MATHEMATICS	4
SCIENCE	5
SOCIAL STUDIES	6
LIT/FINE ARTS	7

④ MAKE YOUR PROBABILITY JUDGMENTS FOR EACH ITEM ON THE FOLLOWING SCALE, FILL-IN THE BUBBLE ON EACH LINE CORRESPONDING TO THESE SCALE POINTS.

1 = .0 THRU .1
2 = .11 THRU .2
3 = .21 THRU .3
4 = .31 THRU .4
5 = .41 THRU .5
6 = .51 THRU .6
7 = .61 THRU .7
8 = .71 THRU .8
9 = .81 THRU .9
10 = .91 THRU 1.0

② FILL-IN APPROPRIATE BUBBLES TO RECORD SOCIAL SECURITY NO. ON LINES BELOW. (MARK ONE BUBBLE ON EACH LINE)

⊂0⊃ ⊂1⊃ ⊂2⊃ ⊂3⊃ ⊂4⊃ ⊂5⊃ ⊂6⊃ ⊂7⊃ ⊂8⊃ ⊂9⊃
⊂0⊃ ⊂1⊃ ⊂2⊃ ⊂3⊃ ⊂4⊃ ⊂5⊃ ⊂6⊃ ⊂7⊃ ⊂8⊃ ⊂9⊃
⊂0⊃ ⊂1⊃ ⊂2⊃ ⊂3⊃ ⊂4⊃ ⊂5⊃ ⊂6⊃ ⊂7⊃ ⊂8⊃ ⊂9⊃
⊂0⊃ ⊂1⊃ ⊂2⊃ ⊂3⊃ ⊂4⊃ ⊂5⊃ ⊂6⊃ ⊂7⊃ ⊂8⊃ ⊂9⊃
⊂0⊃ ⊂1⊃ ⊂2⊃ ⊂3⊃ ⊂4⊃ ⊂5⊃ ⊂6⊃ ⊂7⊃ ⊂8⊃ ⊂9⊃
⊂0⊃ ⊂1⊃ ⊂2⊃ ⊂3⊃ ⊂4⊃ ⊂5⊃ ⊂6⊃ ⊂7⊃ ⊂8⊃ ⊂9⊃
⊂0⊃ ⊂1⊃ ⊂2⊃ ⊂3⊃ ⊂4⊃ ⊂5⊃ ⊂6⊃ ⊂7⊃ ⊂8⊃ ⊂9⊃
⊂0⊃ ⊂1⊃ ⊂2⊃ ⊂3⊃ ⊂4⊃ ⊂5⊃ ⊂6⊃ ⊂7⊃ ⊂8⊃ ⊂9⊃
⊂0⊃ ⊂1⊃ ⊂2⊃ ⊂3⊃ ⊂4⊃ ⊂5⊃ ⊂6⊃ ⊂7⊃ ⊂8⊃ ⊂9⊃

③ RECORD TEST NO. ON LINE BELOW

⊂1⊃ ⊂2⊃ ⊂3⊃ ⊂4⊃ ⊂5⊃ ⊂6⊃ ⊂7⊃

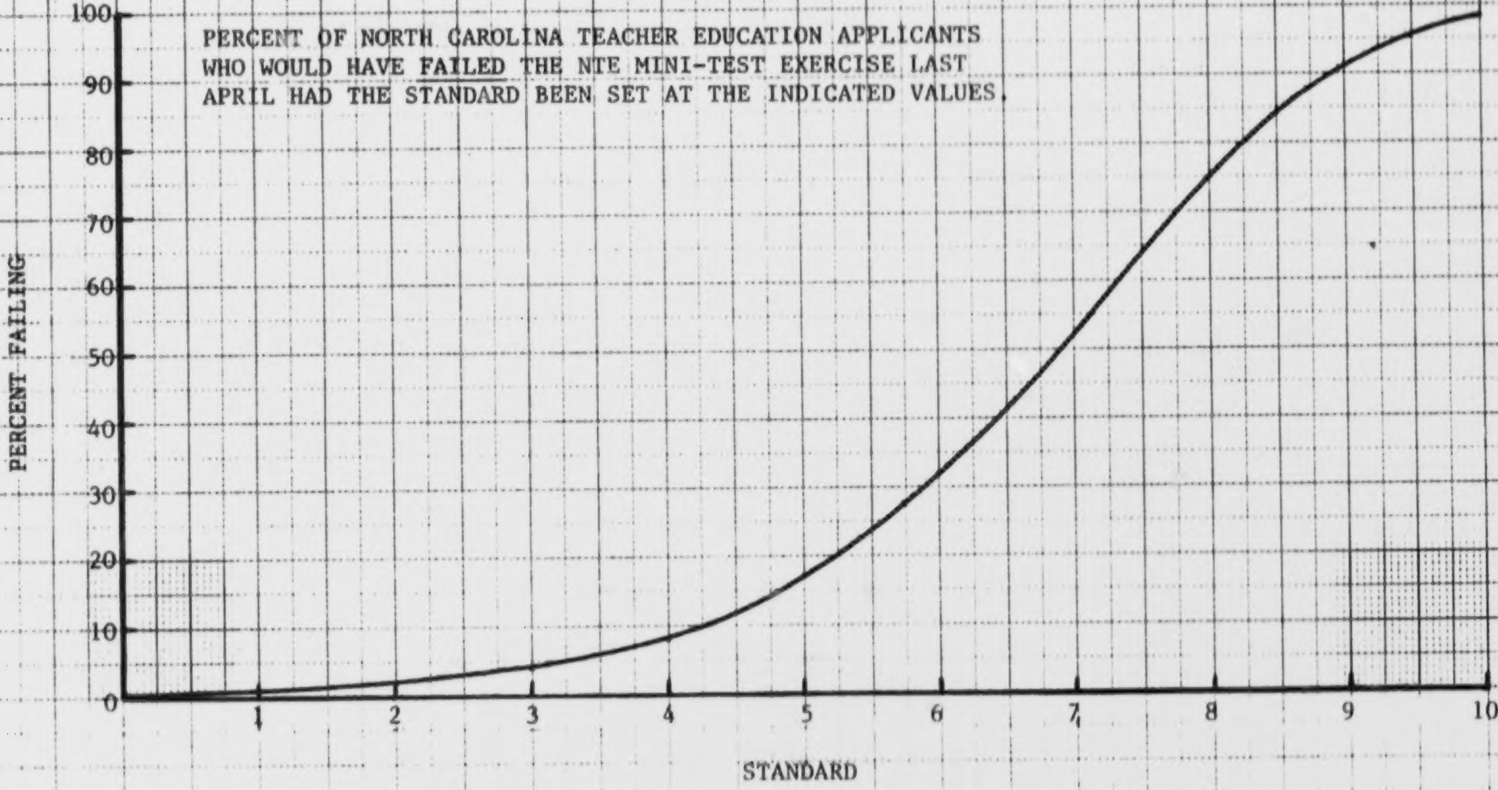
④ RECORD PROBABILITY JUDGMENTS BELOW

ITEM NO. ⊂1⊃ ⊂2⊃ ⊂3⊃ ⊂4⊃ ⊂5⊃ ⊂6⊃ ⊂7⊃ ⊂8⊃ ⊂9⊃ ⊂10⊃

1 ⊂1⊃ ⊂2⊃ ⊂3⊃ ⊂4⊃ ⊂5⊃ ⊂6⊃ ⊂7⊃ ⊂8⊃ ⊂9⊃ ⊂10⊃
2 ⊂1⊃ ⊂2⊃ ⊂3⊃ ⊂4⊃ ⊂5⊃ ⊂6⊃ ⊂7⊃ ⊂8⊃ ⊂9⊃ ⊂10⊃
3 ⊂1⊃ ⊂2⊃ ⊂3⊃ ⊂4⊃ ⊂5⊃ ⊂6⊃ ⊂7⊃ ⊂8⊃ ⊂9⊃ ⊂10⊃
4 ⊂1⊃ ⊂2⊃ ⊂3⊃ ⊂4⊃ ⊂5⊃ ⊂6⊃ ⊂7⊃ ⊂8⊃ ⊂9⊃ ⊂10⊃
5 ⊂1⊃ ⊂2⊃ ⊂3⊃ ⊂4⊃ ⊂5⊃ ⊂6⊃ ⊂7⊃ ⊂8⊃ ⊂9⊃ ⊂10⊃
6 ⊂1⊃ ⊂2⊃ ⊂3⊃ ⊂4⊃ ⊂5⊃ ⊂6⊃ ⊂7⊃ ⊂8⊃ ⊂9⊃ ⊂10⊃
7 ⊂1⊃ ⊂2⊃ ⊂3⊃ ⊂4⊃ ⊂5⊃ ⊂6⊃ ⊂7⊃ ⊂8⊃ ⊂9⊃ ⊂10⊃
8 ⊂1⊃ ⊂2⊃ ⊂3⊃ ⊂4⊃ ⊂5⊃ ⊂6⊃ ⊂7⊃ ⊂8⊃ ⊂9⊃ ⊂10⊃
9 ⊂1⊃ ⊂2⊃ ⊂3⊃ ⊂4⊃ ⊂5⊃ ⊂6⊃ ⊂7⊃ ⊂8⊃ ⊂9⊃ ⊂10⊃
10 ⊂1⊃ ⊂2⊃ ⊂3⊃ ⊂4⊃ ⊂5⊃ ⊂6⊃ ⊂7⊃ ⊂8⊃ ⊂9⊃ ⊂10⊃
11 ⊂1⊃ ⊂2⊃ ⊂3⊃ ⊂4⊃ ⊂5⊃ ⊂6⊃ ⊂7⊃ ⊂8⊃ ⊂9⊃ ⊂10⊃
12 ⊂1⊃ ⊂2⊃ ⊂3⊃ ⊂4⊃ ⊂5⊃ ⊂6⊃ ⊂7⊃ ⊂8⊃ ⊂9⊃ ⊂10⊃
13 ⊂1⊃ ⊂2⊃ ⊂3⊃ ⊂4⊃ ⊂5⊃ ⊂6⊃ ⊂7⊃ ⊂8⊃ ⊂9⊃ ⊂10⊃
14 ⊂1⊃ ⊂2⊃ ⊂3⊃ ⊂4⊃ ⊂5⊃ ⊂6⊃ ⊂7⊃ ⊂8⊃ ⊂9⊃ ⊂10⊃
15 ⊂1⊃ ⊂2⊃ ⊂3⊃ ⊂4⊃ ⊂5⊃ ⊂6⊃ ⊂7⊃ ⊂8⊃ ⊂9⊃ ⊂10⊃
16 ⊂1⊃ ⊂2⊃ ⊂3⊃ ⊂4⊃ ⊂5⊃ ⊂6⊃ ⊂7⊃ ⊂8⊃ ⊂9⊃ ⊂10⊃
17 ⊂1⊃ ⊂2⊃ ⊂3⊃ ⊂4⊃ ⊂5⊃ ⊂6⊃ ⊂7⊃ ⊂8⊃ ⊂9⊃ ⊂10⊃
18 ⊂1⊃ ⊂2⊃ ⊂3⊃ ⊂4⊃ ⊂5⊃ ⊂6⊃ ⊂7⊃ ⊂8⊃ ⊂9⊃ ⊂10⊃
19 ⊂1⊃ ⊂2⊃ ⊂3⊃ ⊂4⊃ ⊂5⊃ ⊂6⊃ ⊂7⊃ ⊂8⊃ ⊂9⊃ ⊂10⊃
20 ⊂1⊃ ⊂2⊃ ⊂3⊃ ⊂4⊃ ⊂5⊃ ⊂6⊃ ⊂7⊃ ⊂8⊃ ⊂9⊃ ⊂10⊃
21 ⊂1⊃ ⊂2⊃ ⊂3⊃ ⊂4⊃ ⊂5⊃ ⊂6⊃ ⊂7⊃ ⊂8⊃ ⊂9⊃ ⊂10⊃
22 ⊂1⊃ ⊂2⊃ ⊂3⊃ ⊂4⊃ ⊂5⊃ ⊂6⊃ ⊂7⊃ ⊂8⊃ ⊂9⊃ ⊂10⊃
23 ⊂1⊃ ⊂2⊃ ⊂3⊃ ⊂4⊃ ⊂5⊃ ⊂6⊃ ⊂7⊃ ⊂8⊃ ⊂9⊃ ⊂10⊃
24 ⊂1⊃ ⊂2⊃ ⊂3⊃ ⊂4⊃ ⊂5⊃ ⊂6⊃ ⊂7⊃ ⊂8⊃ ⊂9⊃ ⊂10⊃
25 ⊂1⊃ ⊂2⊃ ⊂3⊃ ⊂4⊃ ⊂5⊃ ⊂6⊃ ⊂7⊃ ⊂8⊃ ⊂9⊃ ⊂10⊃
26 ⊂1⊃ ⊂2⊃ ⊂3⊃ ⊂4⊃ ⊂5⊃ ⊂6⊃ ⊂7⊃ ⊂8⊃ ⊂9⊃ ⊂10⊃
27 ⊂1⊃ ⊂2⊃ ⊂3⊃ ⊂4⊃ ⊂5⊃ ⊂6⊃ ⊂7⊃ ⊂8⊃ ⊂9⊃ ⊂10⊃
28 ⊂1⊃ ⊂2⊃ ⊂3⊃ ⊂4⊃ ⊂5⊃ ⊂6⊃ ⊂7⊃ ⊂8⊃ ⊂9⊃ ⊂10⊃
29 ⊂1⊃ ⊂2⊃ ⊂3⊃ ⊂4⊃ ⊂5⊃ ⊂6⊃ ⊂7⊃ ⊂8⊃ ⊂9⊃ ⊂10⊃
30 ⊂1⊃ ⊂2⊃ ⊂3⊃ ⊂4⊃ ⊂5⊃ ⊂6⊃ ⊂7⊃ ⊂8⊃ ⊂9⊃ ⊂10⊃
31 ⊂1⊃ ⊂2⊃ ⊂3⊃ ⊂4⊃ ⊂5⊃ ⊂6⊃ ⊂7⊃ ⊂8⊃ ⊂9⊃ ⊂10⊃
32 ⊂1⊃ ⊂2⊃ ⊂3⊃ ⊂4⊃ ⊂5⊃ ⊂6⊃ ⊂7⊃ ⊂8⊃ ⊂9⊃ ⊂10⊃
33 ⊂1⊃ ⊂2⊃ ⊂3⊃ ⊂4⊃ ⊂5⊃ ⊂6⊃ ⊂7⊃ ⊂8⊃ ⊂9⊃ ⊂10⊃
34 ⊂1⊃ ⊂2⊃ ⊂3⊃ ⊂4⊃ ⊂5⊃ ⊂6⊃ ⊂7⊃ ⊂8⊃ ⊂9⊃ ⊂10⊃
35 ⊂1⊃ ⊂2⊃ ⊂3⊃ ⊂4⊃ ⊂5⊃ ⊂6⊃ ⊂7⊃ ⊂8⊃ ⊂9⊃ ⊂10⊃
36 ⊂1⊃ ⊂2⊃ ⊂3⊃ ⊂4⊃ ⊂5⊃ ⊂6⊃ ⊂7⊃ ⊂8⊃ ⊂9⊃ ⊂10⊃
37 ⊂1⊃ ⊂2⊃ ⊂3⊃ ⊂4⊃ ⊂5⊃ ⊂6⊃ ⊂7⊃ ⊂8⊃ ⊂9⊃ ⊂10⊃
38 ⊂1⊃ ⊂2⊃ ⊂3⊃ ⊂4⊃ ⊂5⊃ ⊂6⊃ ⊂7⊃ ⊂8⊃ ⊂9⊃ ⊂10⊃
39 ⊂1⊃ ⊂2⊃ ⊂3⊃ ⊂4⊃ ⊂5⊃ ⊂6⊃ ⊂7⊃ ⊂8⊃ ⊂9⊃ ⊂10⊃
40 ⊂1⊃ ⊂2⊃ ⊂3⊃ ⊂4⊃ ⊂5⊃ ⊂6⊃ ⊂7⊃ ⊂8⊃ ⊂9⊃ ⊂10⊃
41 ⊂1⊃ ⊂2⊃ ⊂3⊃ ⊂4⊃ ⊂5⊃ ⊂6⊃ ⊂7⊃ ⊂8⊃ ⊂9⊃ ⊂10⊃
42 ⊂1⊃ ⊂2⊃ ⊂3⊃ ⊂4⊃ ⊂5⊃ ⊂6⊃ ⊂7⊃ ⊂8⊃ ⊂9⊃ ⊂10⊃
43 ⊂1⊃ ⊂2⊃ ⊂3⊃ ⊂4⊃ ⊂5⊃ ⊂6⊃ ⊂7⊃ ⊂8⊃ ⊂9⊃ ⊂10⊃
44 ⊂1⊃ ⊂2⊃ ⊂3⊃ ⊂4⊃ ⊂5⊃ ⊂6⊃ ⊂7⊃ ⊂8⊃ ⊂9⊃ ⊂10⊃
45 ⊂1⊃ ⊂2⊃ ⊂3⊃ ⊂4⊃ ⊂5⊃ ⊂6⊃ ⊂7⊃ ⊂8⊃ ⊂9⊃ ⊂10⊃

TELEDYNE
NATIONAL TRACING PAPER DIVISION
INDIANAPOLIS INDIANA

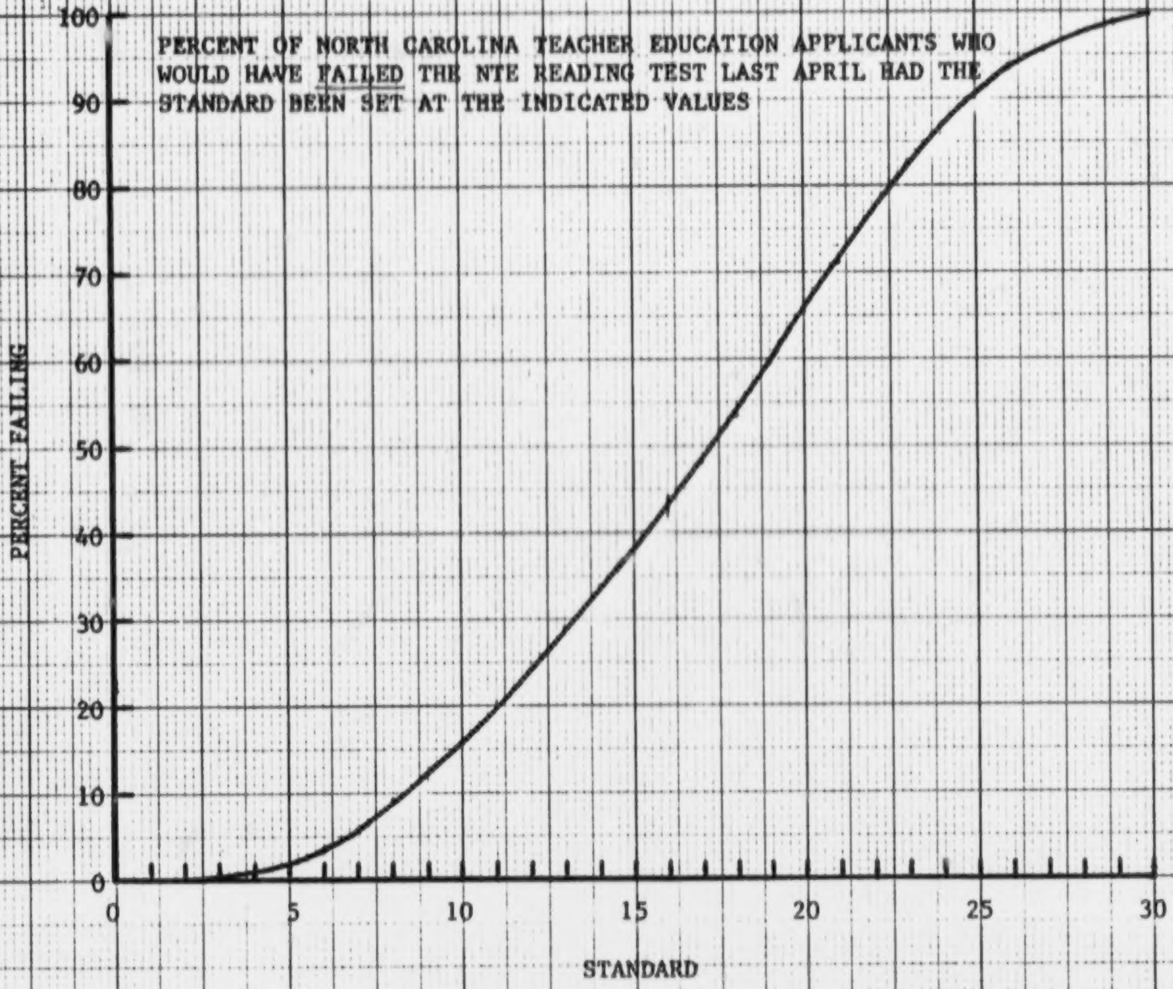
NO. 2301TGM MILLIMETER
LITHO IN U.S.A.



PERCENT OF NORTH CAROLINA TEACHER EDUCATION APPLICANTS WHO WOULD HAVE FAILED THE NTE MINI-TEST EXERCISE LAST APRIL HAD THE STANDARD BEEN SET AT THE INDICATED VALUES.

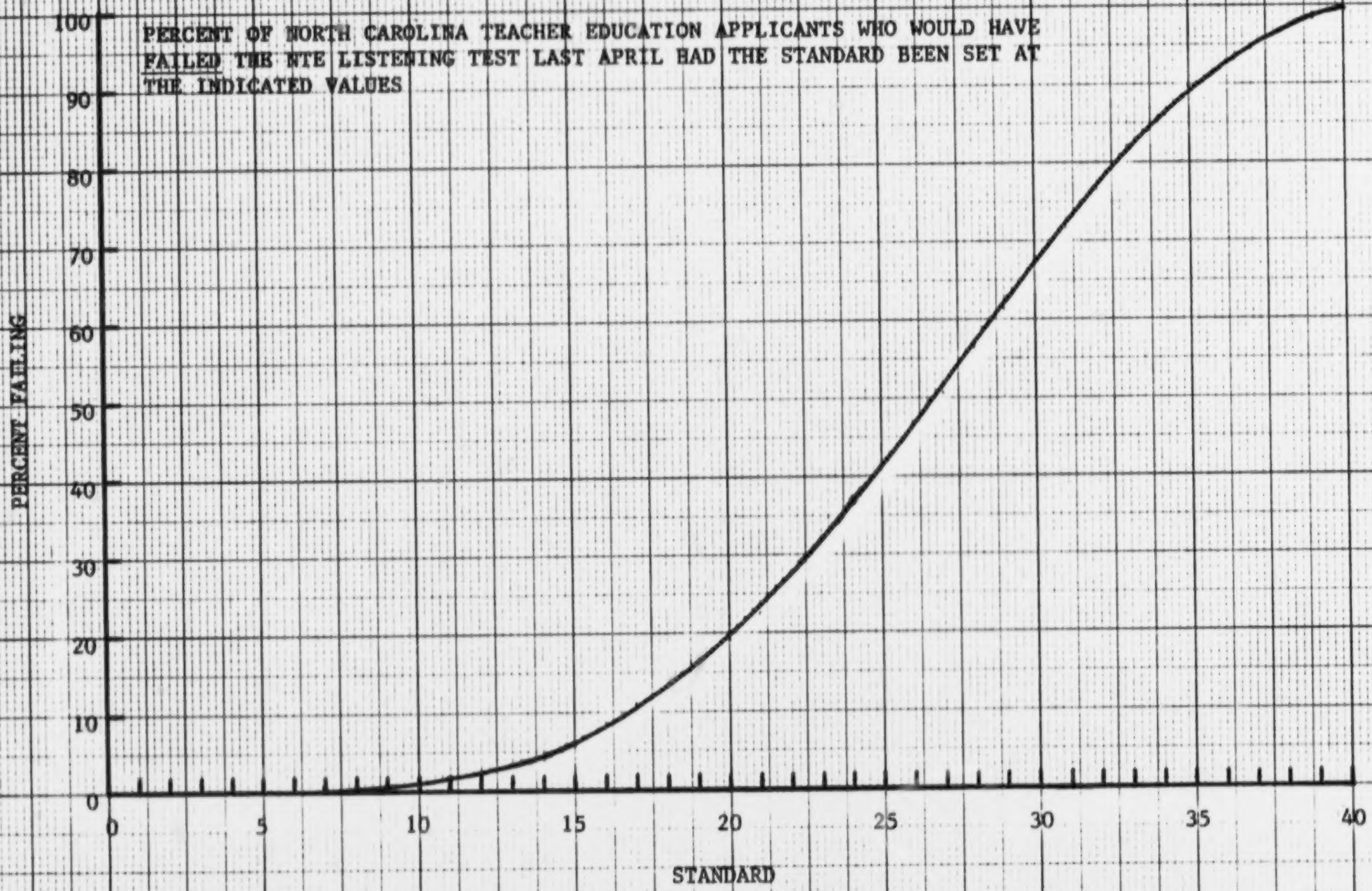
TELEDYNE POST

20MA 500 20X20 PER INCH
LITHO IN U.S.A.



PERCENT OF NORTH CAROLINA TEACHER EDUCATION APPLICANTS WHO WOULD HAVE FAILED THE NTE READING TEST LAST APRIL HAD THE STANDARD BEEN SET AT THE INDICATED VALUES

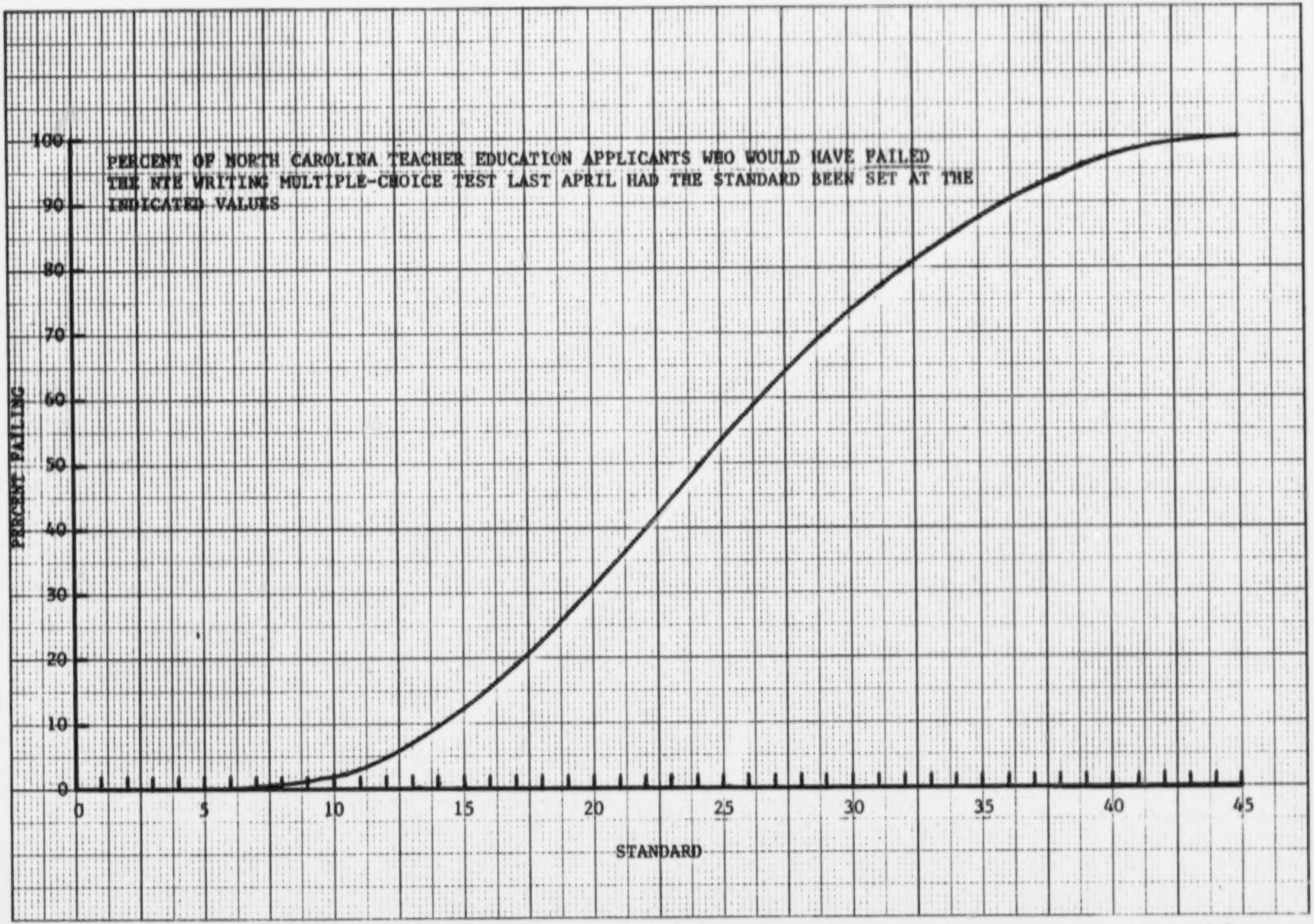
TELEDYNE POST
20MA 500 20X20 PER INCH
LITHO IN U.S.A.



PERCENT OF NORTH CAROLINA TEACHER EDUCATION APPLICANTS WHO WOULD HAVE FAILED THE NTE LISTENING TEST LAST APRIL HAD THE STANDARD BEEN SET AT THE INDICATED VALUES

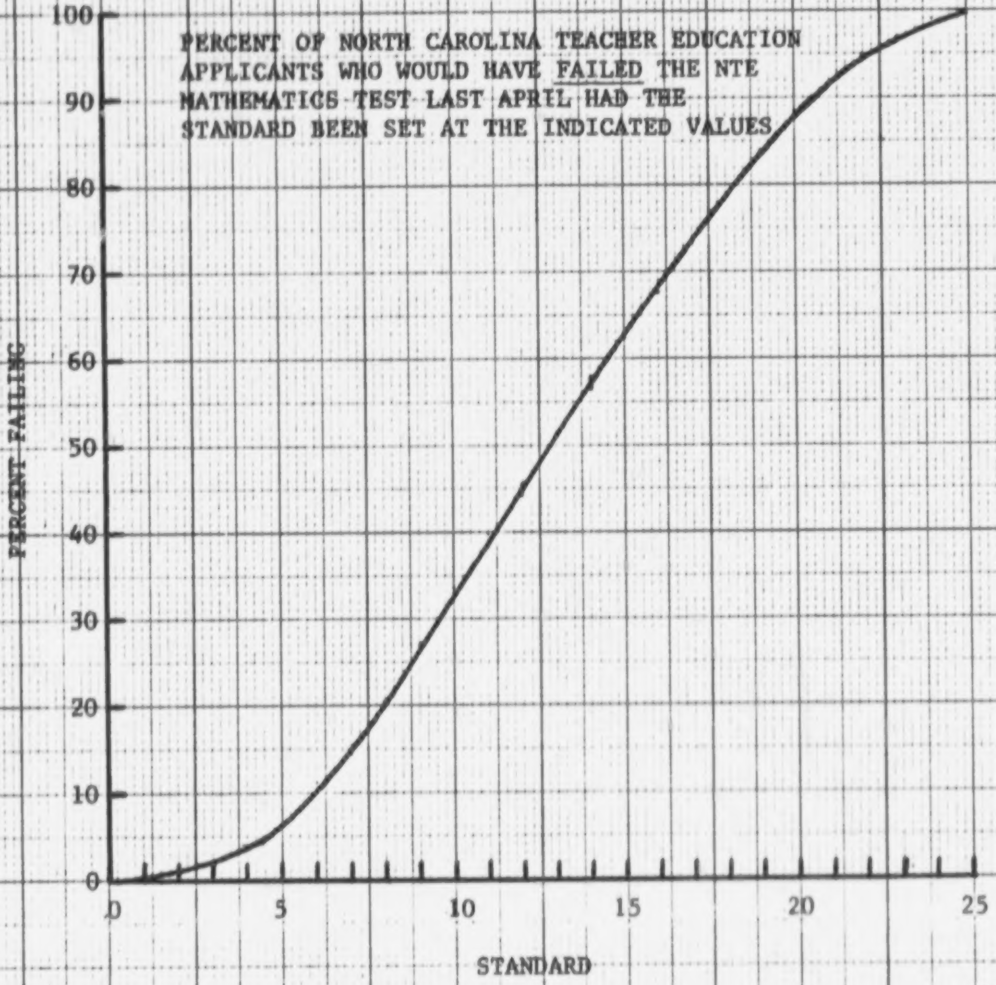
TELEDYNE POST

10MA 500 20X20 PER INCH
LITHO IN U.S.A.



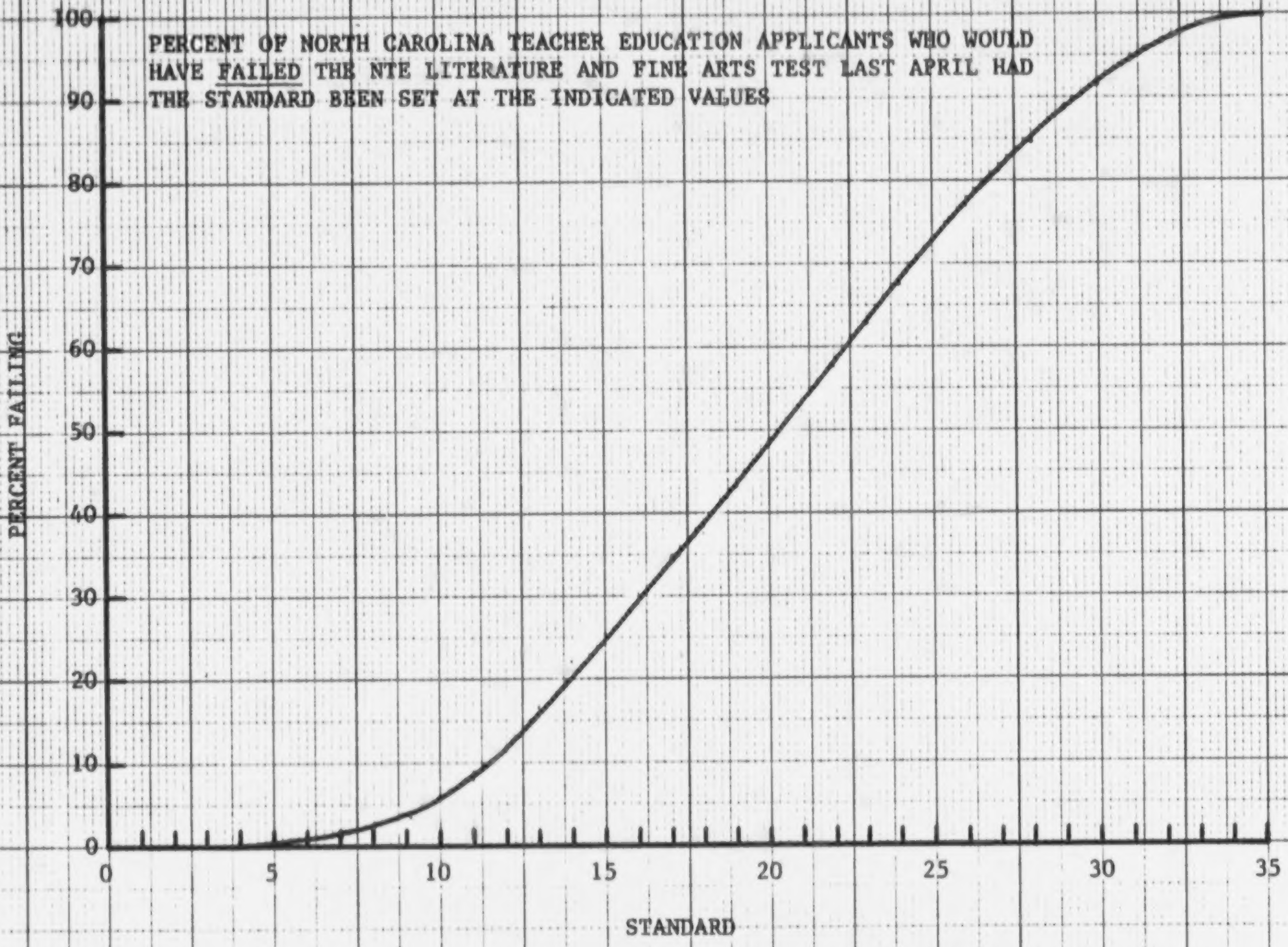
PERCENT OF NORTH CAROLINA TEACHER EDUCATION APPLICANTS WHO WOULD HAVE FAILED THE NTE WRITING MULTIPLE-CHOICE TEST LAST APRIL HAD THE STANDARD BEEN SET AT THE INDICATED VALUES

TELEDYNE POST
20MA-500 20X20 PER INCH
LITHO IN U.S.A.

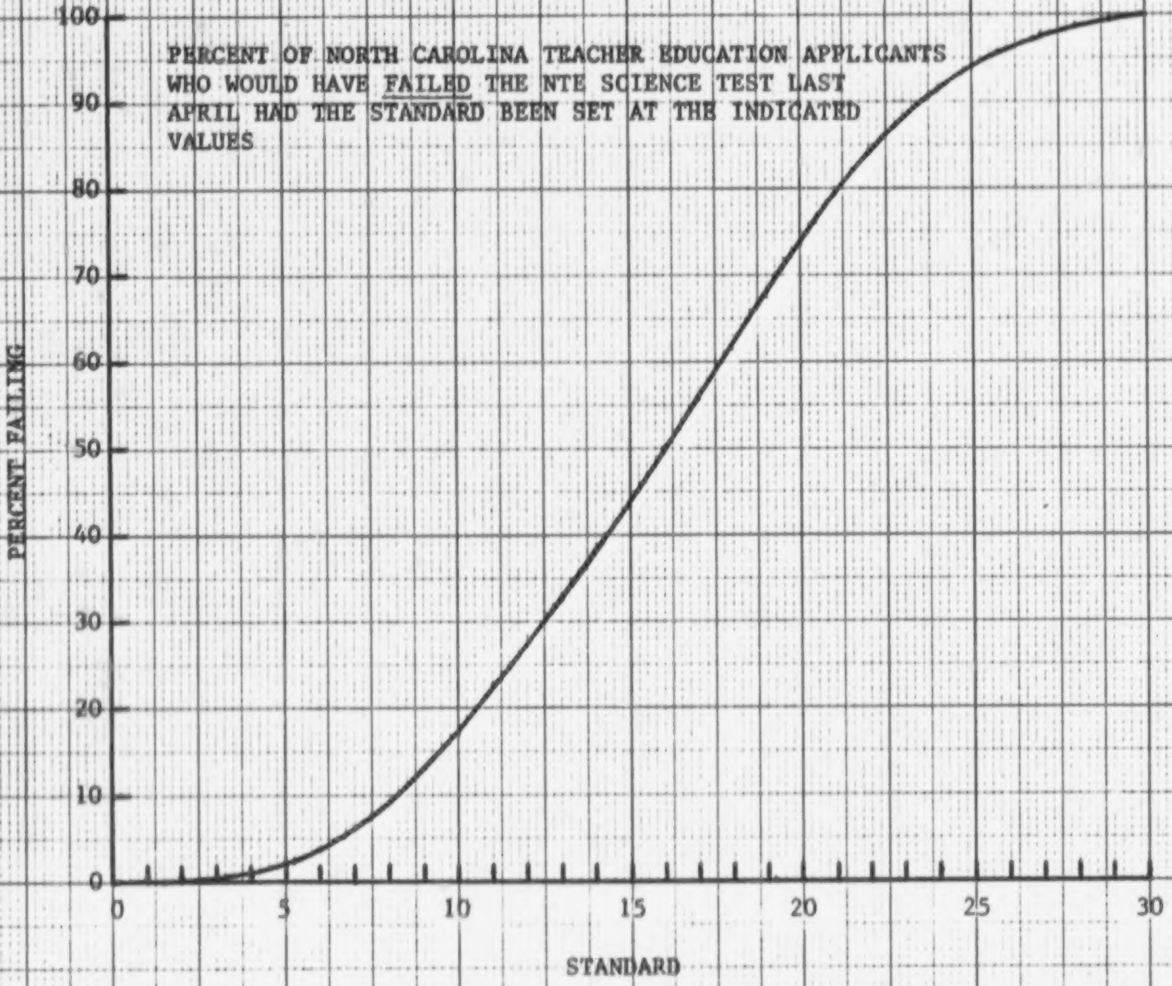


PERCENT OF NORTH CAROLINA TEACHER EDUCATION APPLICANTS WHO WOULD HAVE FAILED THE NTE MATHEMATICS TEST LAST APRIL HAD THE STANDARD BEEN SET AT THE INDICATED VALUES

TELEDYNE POST
20MA 500 20X20 PER INCH
LITHO IN U.S.A.


PERCENT OF NORTH CAROLINA TEACHER EDUCATION APPLICANTS WHO WOULD HAVE FAILED THE NTE LITERATURE AND FINE ARTS TEST LAST APRIL HAD THE STANDARD BEEN SET AT THE INDICATED VALUES

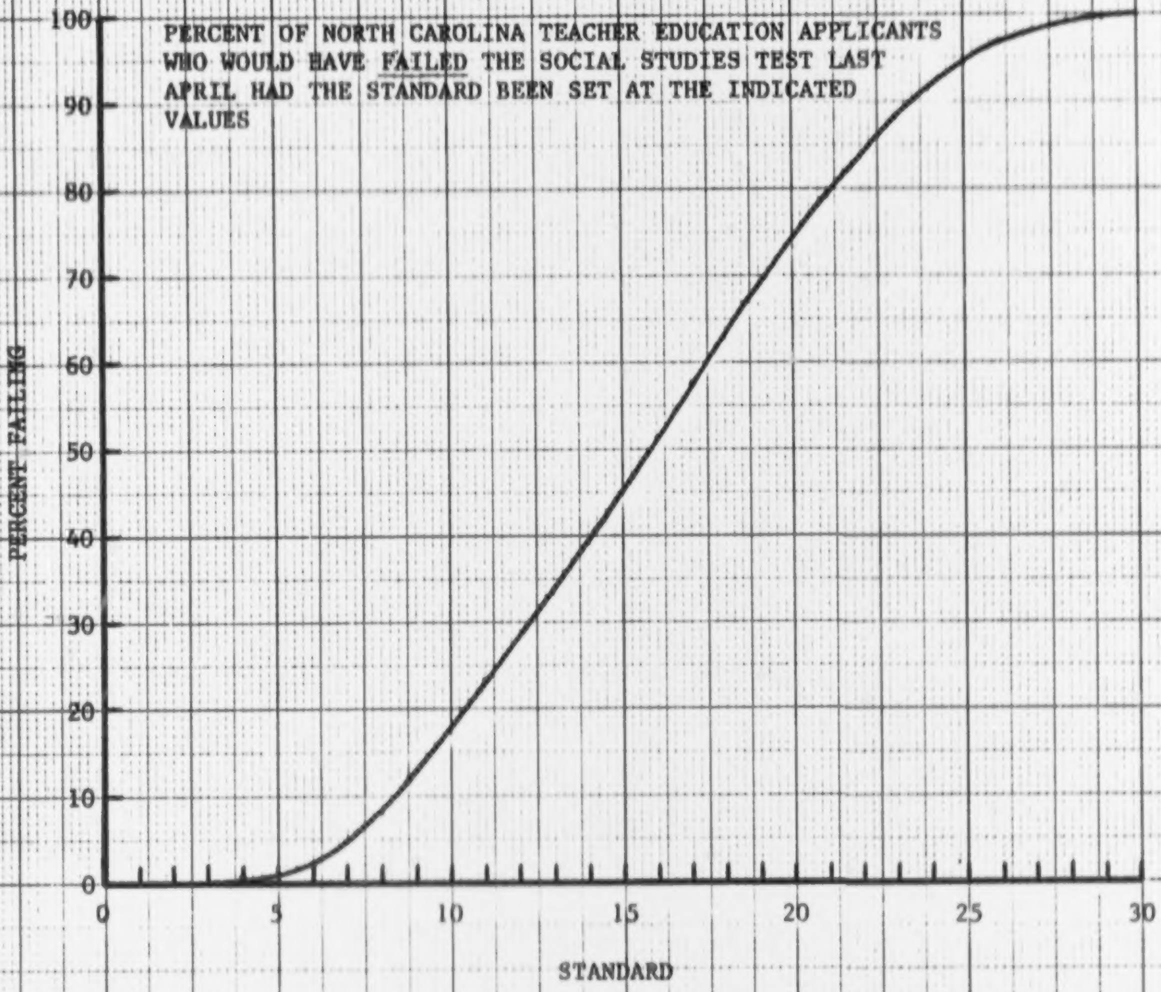
TELEDYNE POST
20MA 500 20X20 PER INCH
LITHO IN U.S.A



PERCENT OF NORTH CAROLINA TEACHER EDUCATION APPLICANTS WHO WOULD HAVE FAILED THE NTE SCIENCE TEST LAST APRIL HAD THE STANDARD BEEN SET AT THE INDICATED VALUES

TELEDYNE POST

20MA 500 20X20 PER INCH
LITHO IN U.S.A.



PERCENT OF NORTH CAROLINA TEACHER EDUCATION APPLICANTS WHO WOULD HAVE FAILED THE SOCIAL STUDIES TEST LAST APRIL HAD THE STANDARD BEEN SET AT THE INDICATED VALUES

```
369121518    12.34      30        30
341282046    12.315     30        30
333571023    13.515     30        30
258558528    11.93      30        30
196032193    13.715     30        30
981110045    15.62      30        30
977542001    10.025     30        30
504278937    16.44      30        30
203292665    13.72      30        30
246810121    11.54      30        30
012345678    1.5        30        30
987654321    28.65      30        30
000000009    15.135     30        30


A TOTAL OF 13   FORMS HAVE BEEN PROCESSED.
ITEM RESPONSES FOR SCIENCE

ITEM 1
 1  ***
 2  **
 3  *
 4  *
 5  *
 6
 7  *
 8
 9
 10 ****
 11
ITEM 2
 1  ****
 2  **
 3
 4  **
 5
 6  **
 7
 8
 9  *
 10 **
 11
ITEM 3
 1  ***
 2  *
 3  **
 4  *
 5
 6  *
 7  *
 8
 9  *
 10 ***
 11
ITEM 4
 1  *****
 2
 3
 4  ***
 5  **
 6  **
 7
 8
```

```
ITEM 5
 1  ***
 2  ****
 3
 4  **
 5  *
 6
 7
 8  *
 9  *
 10 *
 11
ITEM 6
 1  **
 2  *
 3  **
 4
 5
 6  *
 7  **
 8  *
 9  *
 10 ***
 11
ITEM 7
 1  **
 2  **
 3  *
 4
 5  *
 6
 7  **
 8  *
 9  **
 10 **
 11
ITEM 8
 1  **
 2  **
 3  *
 4
 5  *
 6
 7  *
 8  **
 9  *
 10 ***
 11
ITEM 9
 1  **
 2  ***
 3  **
 4
 5  *
 6
 7
 8
 9  ***
 10 **
 11
ITEM 10
 1  **
 2  **
```

GLOBAL EVALUATION OF INSTRUCTIONAL VALIDITY

Judge Name ____________________ Test ____________________

Institution ____________________

Directions: Please describe the degree of similarity between this NTE test and the relevant part of the general education curriculum at your institution which is assessed by this test.

CHECK ONE CATEGORY

________A. The content of this test and the content included in relevant portions of my institution's general education curriculum are very similiar.

________B. There are some differences between the content of this test and the content included in relevant portions of my institution's general education curriculum, but these differences do not appear to be appreciable.

________C. There appear to be some appreciable differences between the content of this test and the content included in relevant portions of my institution's general education curriculum.

________D. There appears to be little similarity between the content of this test and the content included in relevant portions of my institution's general education curriculum.

Name________________________
Institution________________

Instructional Validity Response Sheet for the Literature/Fine Arts Test

Directions: Read each part of the test blueprint and then answer the question below.

"Would 90% or more of the applicants to the teacher education programs at my institution have had a reasonable opportunity to secure instruction in the skills or knowledge defined by each statement in the blueprint prior to application for admission to teacher education?"

Topic From Test Blueprint	Please circle your response to the question.	
1. The ability to recognize basic elements and components of works of literature and fine arts, such as the ability to: Demonstrate knowledge and awareness of basic elements of meaning such as vocabulary in context, basic musical concepts, widely used symbols, representational features, etc. Demonstrate knowledge and awareness of basic structural components and materials including the nature of materials, compositional elements, principles of sound and light, etc.	Yes	No
2. The ability to analyze and interpret works of literature and fine art, such as; Demonstrates understanding of formal structures in terms of the interrelationship of their elements including concepts such as relation of parts of works to the whole, pattern, unity, function, rhythm, etc.	Yes	No

Instructional Validity Response Sheet for the Literature/Fine Arts Test (Continued)

Topic From Test Blueprint	Please circle your response to the question.	
3. The ability to relate works of literature and fine arts to one another and to their social/historical context, such as the ability to: See works of literature and fine arts as reflections of their social/historical context or as embodiments of significant questions concerning human values. Demonstrate awareness of similarities and differences between works from different cultural traditions and historical periods. Distinguish between works from different genres, stylistic types, and schools and seeing relationships between the various arts.	Yes	No

NORTH CAROLINA QUALITY ASSURANCE PROJECT

PARTICIPANT'S EVALUATION QUESTIONNAIRE

Please take a few a moments to give us your judgments on the procedures we have asked you to use today, and to provide some information on your consideration and use of data provided in the judgment sessions.

Your responses will be held in strict confidence, and will not be reported individually. We are asking you to provide your name and social security number in the blanks provided at the top of the RESPONSE FORM only for data analysis purposes.

Notice that the computer-readable response sheet has five columns, headed "SD," "D," "DK," "A," and "SA." These column headings are defined as follows:

SD -- Strongly Disagree
D -- Disagree
DK -- Don't Know
A -- Agree
SA -- Strongly Agree

Please darken the bubble in the appropriate column to indicate your judgments regarding each of the following numbered statements. Be sure to match the number of the statement on this sheet with the number of the row on the Response Sheet.

1. I understand the need to validate the NTE tests for use with teacher education applicants.

2. I do not understand the need to set standards on the NTE tests for use with teacher education applicants.

3. I think it is important to determine the relevance of the NTE tests to success in a North Carolina teacher education program.

4. I do not think it is important to screen candidates for admission to North Carolina teacher education programs.

5. I think that standardized testing is an appropriate way to screen candidates for admission to North Carolina teacher education programs.

6. I do not think that the NTE tests are appropriate for screening candidates for admission to North Carolina teacher education programs.

7. I had all the time I needed to complete the NTE test that I took. NOTE: IF YOU REVIEWED THE WRITING ESSAY TEST, MARK "DK" FOR THIS ITEM.

8. I did not fully understand the directions for providing relevance judgments.

9. I found it very difficult to make judgments on the relevance of NTE test items for success in a teacher education program. NOTE: IF YOU REVIEWED THE WRITING ESSAY TEST, INTERPRET THIS QUESTION AS THOUGH IT APPLIED TO THE ENTIRE ESSAY TEST.

10. I feel confident about my judgments on the relevance of NTE test items for success in a teacher education program. NOTE: IF YOU REVIEWED THE WRITING ESSAY TEST, INTERPRET THIS QUESTION AS THOUGH IT APPLIED TO THE ENTIRE ESSAY TEST.

11. I had enough time to make to make my judgments on the relevance of NTE test items for success in a teacher education program. NOTE: IF YOU REVIEWED THE WRITING ESSAY TEST, INTERPRET THIS QUESTION AS THOUGH IT APPLIED TO THE ENTIRE ESSAY TEST.

12. I understand the meaning of the ten-point probability scale used in the standard-setting procedure. NOTE: IF YOU REVIEWED THE WRITING ESSAY TEST, MARK "DK" FOR THIS ITEM.

13. I do not feel confident about my first (before lunch) estimates of probabilities of correct answers for minimally competent applicants to teacher education programs. NOTE: IF YOU REVIEWED THE WRITING ESSAY TEST, MARK "DK" FOR THIS ITEM.

14. The mini-test exercise provided information that I used in formulating my second set of estimated probabilities for minimally competent applicants. NOTE: IF YOU REVIEWED THE WRITING ESSAY TEST, MARK "DK" FOR THIS ITEM.

15. I did not understand the meaning of the percents in parentheses beside the items on the Mini-Test Exercise. NOTE: IF YOU REVIEWED THE WRITING ESSAY TEST, MARK "DK" FOR THIS ITEM.

16. I fully understood the meaning of the numbers in boxes beside the items on the Mini-Test Exercise. NOTE: IF YOU REVIEWED THE WRITING ESSAY TEST, MARK "DK" FOR THIS ITEM.

17. The graph distributed at the beginning of the second standard-setting session made sense to me.

18. I did not change my estimated probabilities very much from the first standard-setting session to the second one. NOTE: IF YOU REVIEWED THE WRITING ESSAY TEST, MARK "DK" FOR THIS ITEM.

19. My estimated probabilities were lower in the second standard-setting session than in the first one. NOTE: IF YOU REVIEWED THE WRITING ESSAY TEST, MARK "DK" FOR THIS ITEM.

20. Data on the percentage of all applicants who answered each test item correctly did not influence my estimated probabilities during the second standard-setting session. NOTE: IF YOU REVIEWED THE WRITING ESSAY TEST, MARK "DK" FOR THIS ITEM.

21. Data on the percentage of applicants with scores in the 30 to 50-percent range who answered each test item correctly influenced my estimated probabilities during the second standard-setting session. NOTE: IF YOU REVIEWED THE WRITING ESSAY TEST, MARK "DK" FOR THIS ITEM.

22. Data in the graph I received at the beginning of the second standard-setting session, on the percentage of North Carolina applicants who would fail the test, did not influence my estimates during that session.

23. My recommendations during the second standard-setting session were influenced by the discussion of items during that session. NOTE: IF YOU REVIEWED THE WRITING ESSAY TEST, MARK "DK" FOR THIS ITEM.

24. I felt that the instructional validity procedures were explained clearly and completely. NOTE: MARK THIS ITEM "DK" IF YOU ARE A PUBLIC SCHOOL TEACHER.

25. I am confident about my ratings of the instructional validity of the NTE tests. NOTE: MARK THIS ITEM "DK" IF YOU ARE A PUBLIC SCHOOL TEACHER.

PLEASE ELABORATE ON YOUR RESPONSE TO ANY OF THESE QUESTIONS OR MAKE ANY COMMENTS OR SUGGESTIONS ON OUR PROCEDURES IN THE BLANK SPACE TO THE RIGHT OF THE BUBBLES AND THE GREEN VERTICAL LINE ON THE RESPONSE SHEET.

THANK YOU VERY MUCH FOR TAKING THE TIME TO RESPOND
TO THIS QUESTIONNAIRE
YOUR JUDGMENTS WILL HELP US TO IMPROVE OUR PROCEDURES

(T) (F) KEY ( )

	SD	D	DK	A	SA
1	(A)	(B)	(C)	(D)	(E)
2	(A)	(B)	(C)	(D)	(E)
3	(A)	(B)	(C)	(D)	(E)
4	(A)	(B)	(C)	(D)	(E)
5	(A)	(B)	(C)	(D)	(E)
6	(A)	(B)	(C)	(D)	(E)
7	(A)	(B)	(C)	(D)	(E)
8	(A)	(B)	(C)	(D)	(E)
9	(A)	(B)	(C)	(D)	(E)
10	(A)	(B)	(C)	(D)	(E)
11	(A)	(B)	(C)	(D)	(E)
12	(A)	(B)	(C)	(D)	(E)
13	(A)	(B)	(C)	(D)	(E)
14	(A)	(B)	(C)	(D)	(E)
15	(A)	(B)	(C)	(D)	(E)
16	(A)	(B)	(C)	(D)	(E)
17	(A)	(B)	(C)	(D)	(E)
18	(A)	(B)	(C)	(D)	(E)
19	(A)	(B)	(C)	(D)	(E)
20	(A)	(B)	(C)	(D)	(E)
21	(A)	(B)	(C)	(D)	(E)
22	(A)	(B)	(C)	(D)	(E)
23	(A)	(B)	(C)	(D)	(E)
24	(A)	(B)	(C)	(D)	(E)
25	(A)	(B)	(C)	(D)	(E)

↑ FEED THIS DIRECTION

PARTICIPANT'S EVALUATION QUESTIONNAIRE--RESPONSE FORM

NAME: ____

SOCIAL SECURITY NUMBER: ______________________________

COMMENTS OR SUGGESTIONS:

SCAN-TRON™ FORM 899

180

INFORMATION SHEET

For Public School Teachers

NAME: ____

ADDRESS: ____

____

SOCIAL SECURITY NUMBER: ____________________________

TITLE: ____

SCHOOL: ____

SUBJECTS TAUGHT: ___________________________________

TEACHER EDUCATION PROGRAM: __________________________

CERTIFICATION AREA: _________________________________

INFORMATION SHEET

For College and University Faculty

NAME: ______________________________

ADDRESS: ______________________________

____

SOCIAL SECURITY NUMBER: ______________________________

TITLE: ______________________________

INSTITUTION: ______________________________

DEPARTMENT: ______________________________

NAME OF DEPARTMENT HEAD OR DEAN: ______________________________

____

SPECIAL RESPONSIBILITIES: (e.g., teach freshman composition courses; teach English education courses) ______________________________

____

TODAY'S DATE: ______________________________

ETS SPECIFICATIONS FOR ESSAY TEST

This statement summarizes the view of the Educational Testing Service Writing Subcommittee on what an effective writer should be able to do and what the teaching of effective writing should emphasize. These abilities include:

1. provide and sustain a focus or thesis
2. attain in different papers the varied aims or purposes (e.g., explanatory, persuasive, expressive) of discourse
3. decide which of these aims or purposes is appropriate in a given writing situation
4. select and sustain an appropriate persona or voice
5. produce and develop adequate and appropriate material to accomplish the purpose for writing, identifying and supporting, as appropriate, important assumptions
6. choose and use a mode of organization consistently
7. preserve coherence in an extended piece of writing
8. choose an appropriate mode of organization (chronological, enumerative)
9. construct sentences in standard written English, adjusting choice of sentence structure and word choice to suit purposes and aims
10. use sentences and vocabulary which are appropriate to the purpose of the writing
11. use words and sentences which are appropriate for the intended readers
12. construct sentences in standard written English and identify sentences that do not observe the conventions of standard written English, such as grammar, usage, and punctuation

NORTH CAROLINA TEACHER QUALITY ASSURANCE PROGRAM

ESSAY TEST STANDARD SETTING FORM

NAME ____________________

EMPLOYED AT (COLLEGE OR SCHOOL) ____________________ DATE ____________

ESSAY SET #1

Please darken the bubble under A if you judge the essay to be "acceptable."
Please darken the bubble under U if you judge the essay to be "unacceptable."

(T)(F)

				A	U	ESSAY
1	(A)	(B)	(C)	(D)	(E)	Dogwood
2	(A)	(B)	(C)	(D)	(E)	Oak
3	(A)	(B)	(C)	(D)	(E)	Fir
4	(A)	(B)	(C)	(D)	(E)	Birch
5	(A)	(B)	(C)	(D)	(E)	Hickory
6	(A)	(B)	(C)	(D)	(E)	Walnut
7	(A)	(B)	(C)	(D)	(E)	Maple
8	(A)	(B)	(C)	(D)	(E)	Pine
9	(A)	(B)	(C)	(D)	(E)	Spruce
10	(A)	(B)	(C)	(D)	(E)	Mahogany
11	(A)	(B)	(C)	(D)	(E)	Locust
12	(A)	(B)	(C)	(D)	(E)	
13	(A)	(B)	(C)	(D)	(E)	
14	(A)	(B)	(C)	(D)	(E)	
15	(A)	(B)	(C)	(D)	(E)	
16	(A)	(B)	(C)	(D)	(E)	
17	(A)	(B)	(C)	(D)	(E)	
18	(A)	(B)	(C)	(D)	(E)	
19	(A)	(B)	(C)	(D)	(E)	
20	(A)	(B)	(C)	(D)	(E)	
21	(A)	(B)	(C)	(D)	(E)	
22	(A)	(B)	(C)	(D)	(E)	
23	(A)	(B)	(C)	(D)	(E)	
24	(A)	(B)	(C)	(D)	(E)	
25	(A)	(B)	(C)	(D)	(E)	

FEED THIS DIRECTION

SCAN-TRON FORM 899

NORTH CAROLINA TEACHER QUALITY ASSURANCE PROGRAM

ESSAY TEST STANDARD SETTING FORM

NAME ______________________ EMPLOYED AT (COLLEGE OR SCHOOL) ______________________ DATE ____________

ESSAY SET #2

Please darken the bubble under A if you judge the essay to be "acceptable."
Please darken the bubble under U if you judge the essay to be "unacceptable."

(T)(F)

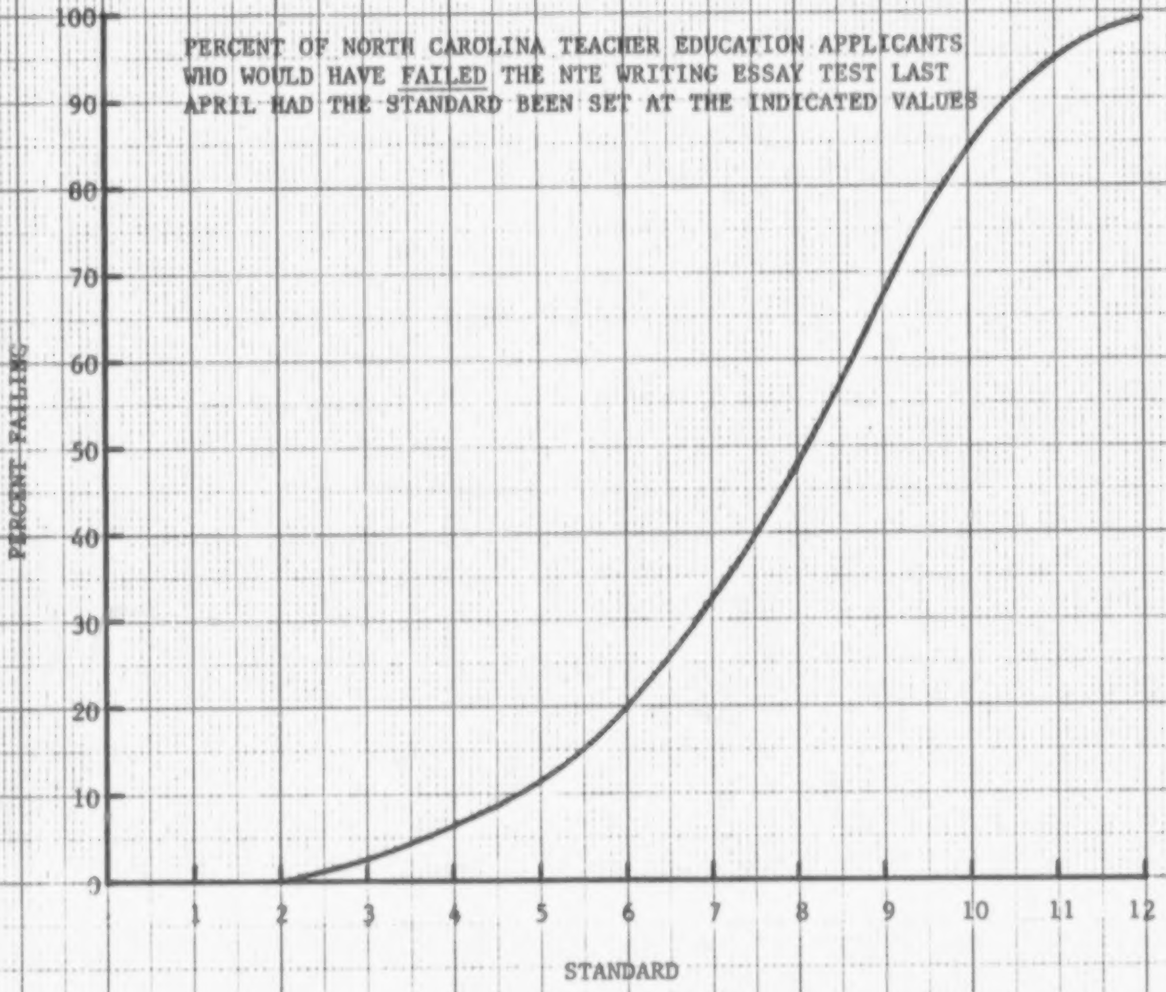
				A	U	ESSAY
1	A	B	C	D	E	Coffee
2	A	B	C	D	E	Willow
3	A	B	C	D	E	Beech
4	A	B	C	D	E	Apple
5	A	B	C	D	E	Juniper
6	A	B	C	D	E	Mesquite
7	A	B	C	D	E	Tamarack
8	A	B	C	D	E	Cherry
9	A	B	C	D	E	Redwood
10	A	B	C	D	E	Orange
11	A	B	C	D	E	Avocado
12	A	B	C	D	E	
13	A	B	C	D	E	
14	A	B	C	D	E	
15	A	B	C	D	E	
16	A	B	C	D	E	
17	A	B	C	D	E	
18	A	B	C	D	E	
19	A	B	C	D	E	
20	A	B	C	D	E	
21	A	B	C	D	E	
22	A	B	C	D	E	
23	A	B	C	D	E	
24	A	B	C	D	E	
25	A	B	C	D	E	

FEED THIS DIRECTION

SCAN-TRON FORM 899

TELEDYNE POST

20MA 500 20X20 PER INCH
LITHO IN U.S.A.



ESSAY RATING FORM -- SECOND JUDGMENT SESSION

For Essay Set No. 1

Name ______________________________

Social Security Number ____________________

School, College, or University ________________________

ESSAY NAME	SCORE GIVEN BY ETS READERS
DOGWOOD	2
OAK	3
FIR	4
BIRCH	5
HICKORY	6
WALNUT	7
MAPLE	8
PINE	9
SPRUCE	10
MAHOGANY	11
LOCUST	12

Write the ETS scores of the **ESSAYS YOU JUDGED TO BE ACCEPTABLE** in the spaces below (use as many spaces as you need):

___ ___ ___ ___ ___ ___ ___ ___ ___ ___ ___ ___

Find the **ACCEPTABLE** essay with the **LOWEST** ETS SCORE. Circle its number in the following list:

2 3 4 5 6 7 8 9 10 11 12

Write the ETS scores of the **ESSAYS YOU JUDGED TO BE UNACCEPTABLE** in the spaces below (use as many spaces as you need):

___ ___ ___ ___ ___ ___ ___ ___ ___ ___ ___ ___

Find the **UNACCEPTABLE** essay with the **HIGHEST** ETS SCORE. Circle its number in the following list:

2 3 4 5 6 7 8 9 10 11 12

ESSAY RATING FORM -- SECOND JUDGMENT SESSION

For Essay Set No. 2

Name ______________________________________

Social Security Number ______________________

School, College, or University ___________________________

ESSAY NAME	SCORE GIVEN BY ETS READERS
COFFEE	2
WILLOW	3
BEECH	4
APPLE	5
JUNIPER	6
MESQUITE	7
TAMARACK	8
CHERRY	9
REDWOOD	10
ORANGE	11
AVOCADO	12

Write the ETS scores of the **ESSAYS YOU JUDGED TO BE ACCEPTABLE** in the spaces below (use as many spaces as you need):

___ ___ ___ ___ ___ ___ ___ ___ ___ ___ ___ ___

Find the **ACCEPTABLE** essay with the **LOWEST** ETS SCORE. Circle its number in the following list:

2 3 4 5 6 7 8 9 10 11 12

Write the ETS scores of the **ESSAYS YOU JUDGED TO BE UNACCEPTABLE** in the spaces below (use as many spaces as you need):

___ ___ ___ ___ ___ ___ ___ ___ ___ ___ ___ ___

Find the **UNACCEPTABLE** essay with the **HIGHEST** ETS SCORE. Circle its number in the following list:

2 3 4 5 6 7 8 9 10 11 12

ESSAY RATING FORM -- THIRD JUDGMENT SESSION

For Essay Set No. 1

Name ______________________________________

Social Security Number ______________________

School, College, or University ____________________________

ESSAY NAME	SCORE GIVEN BY ETS READERS
DOGWOOD	2
OAK	3
FIR	4
BIRCH	5
HICKORY	6
WALNUT	7
MAPLE	8
PINE	9
SPRUCE	10
MAHOGANY	11
LOCUST	12

Write the ETS scores of the **ESSAYS YOU JUDGED TO BE ACCEPTABLE** in the spaces below (use as many spaces as you need):

___ ___ ___ ___ ___ ___ ___ ___ ___ ___ ___ ___

Find the **ACCEPTABLE** essay with the **LOWEST** ETS SCORE. Circle its number in the following list:

2 3 4 5 6 7 8 9 10 11 12

Write the ETS scores of the **ESSAYS YOU JUDGED TO BE UNACCEPTABLE** in the spaces below (use as many spaces as you need):

___ ___ ___ ___ ___ ___ ___ ___ ___ ___ ___ ___

Find the **UNACCEPTABLE** essay with the **HIGHEST** ETS SCORE. Circle its number in the following list:

2 3 4 5 6 7 8 9 10 11 12

ESSAY RATING FORM -- THIRD JUDGMENT SESSION

For Essay Set No. 2

Name ____

Social Security Number ______________________

School, College, or University ____________________________

ESSAY NAME	SCORE GIVEN BY ETS READERS
COFFEE	2
WILLOW	3
BEECH	4
APPLE	5
JUNIPER	6
MESQUITE	7
TAMARACK	8
CHERRY	9
REDWOOD	10
ORANGE	11
AVOCADO	12

Write the ETS scores of the **ESSAYS YOU JUDGED TO BE ACCEPTABLE** in the spaces below (use as many spaces as you need):

___ ___ ___ ___ ___ ___ ___ ___ ___ ___ ___ ___

Find the **ACCEPTABLE** essay with the **LOWEST** ETS SCORE. Circle its number in the following list:

2 3 4 5 6 7 8 9 10 11 12

Write the ETS scores of the **ESSAYS YOU JUDGED TO BE UNACCEPTABLE** in the spaces below (use as many spaces as you need):

___ ___ ___ ___ ___ ___ ___ ___ ___ ___ ___ ___

Find the **UNACCEPTABLE** essay with the **HIGHEST** ETS SCORE. Circle its number in the following list:

2 3 4 5 6 7 8 9 10 11 12

ESSAY TEST CONTENT/RELEVANCE JUDGMENT RESPONSE FORM 1

Name ____

Institution ________________________________

Question 1

Please read the following statement and indicate whether you Strongly Agree (SA), Agree (A), are Undecided (U), Disagree (D), or Strongly Disagree (SD):

"The NTE Essay Examination which I have reviewed, including its administration and scoring procedures, is a satisfactory measure of general writing skill."

Please circle your response:

SA A U D SD

ESSAY TEST CONTENT/RELEVANCE JUDGMENT RESPONSE FORM 2

Name ____

Institution ____________________________________

Question 2

Please consider the writing skills measured by the NTE Essay Test and provide us with your judgment of how important those skills are to success in the teacher education program with which you are most familiar. Specifically, we would like to know how important it would be for a student to have minimal writing skills, as measured by the NTE Essay Test.

Please respond by marking an "X" in the blank in front of one of the following choices:

_____ "Crucial" -- of critical importance to success in a teacher education program. These are skills or knowledge which a student would certainly need in order to be successful in the teacher education curriculum with which you are most familiar.

_____ "Important" -- of great value to success in a teacher education program. These include skills or knowledge which, if possessed by a student, would certainly contribute to a successful experience in teacher education; however it would be possible for a student without some of these skills and knowledge to succeed in teacher education.

_____ "Of Questionable Importance" -- these include skills or knowledge which might enhance performance in teacher education but which are not essential to success.

_____ "Irrelevant to Success in Teacher Education" -- have no value or effect; these skills or knowledge have no bearing whatsoever on successful performance in teacher education.

In the space below, and on the reverse side, please elaborate on the response which you marked above. That is, describe the reasons underlying your choice of a particular response:

ESSAY TEST INSTRUCTIONAL VALIDITY RESPONSE FORM 1

Name ______________________________________

Institution ________________________________

Question 1

Please read the following question and indicate whether you Strongly Agree (SA), Agree (A), are Uncertain (U), Disagree (D), or Strongly Disagree (SD):

"At least 90% of the students applying for admission to teacher education at the institution with which I am most familiar have had a reasonable opportunity to receive instruction in the writing skills described in the ETS statement before applying for admission to teacher education."

PLEASE CIRCLE YOUR RESPONSE

SA A U D SD

In the space below, and on the back of this form, please describe the reasons for your response:

ESSAY TEST INSTRUCTIONAL VALIDITY RESPONSE FORM 2

Name ______________________________________

Institution ________________________________

Question 2

Please read the following statement and indicate whether you Strongly Agree (SA), Agree (A), are Undecided (U), Disagree (D), or Strongly Disagree (SD):

"At least 90% of the students applying for admission to teacher education at the institution with which I am most familiar have had a reasonable opportunity to receive instruction in the writing skills that are actually measured by the NTE Essay test."

PLEASE CIRCLE YOUR RESPONSE

SA A U D SD

In the space below, and on the back of this form, please describe the reasons for your response:

OBSERVATION FORMS FOR GROUP ASSISTANTS -- ALL SESSIONS

Date ______________________________
Location _________________________
Leader ___________________________
# Participants __________________

I INTRODUCTION & BRIEFING (10:00 - 10:20)

Start time __________ Stop time __________

________ Started on time.
If not note reason:
If participants entered late, note entry point in terms of a page number on the script:

________ Leader followed script.
If not, note any deviations:

________ Mechanical procedures went smoothly.
If problems, note the nature of the problems & where they occurred in the script.

________ Participants attention focused on the task.
If not, note point(s) on the script where we "lose" people:

________ Stimulation exercise was successful.
If not, note problems:

________ Leader fielded questions easily.
Note all questions; identify by an (*) if there were any problem questions for the leader.

________ Stopped within time schedule.
If not, note reason:

________ Additional comments on back.

Date ______________________________
Location ___________________________
Leader ____________________________
# Participants ____________________

II TEST TAKEN BY JUDGES (10:20 - 10:50)

Start time __________ Stop time __________

________ Leader followed script.
If not, note any deviations:

________ Mechanical procedures went smoothly:
If problems, note the nature of the problems and where they occured.

________ Participants' attention focussed on task.
If not, note points on the script where we lose people.

________ Leader fielded questions easily.
Note all questions; identify by an (*) if there were any problem questions for the leader.

________ Participants did not discuss test.
Note any deviations:

________ Materials were not taken from the testing room.
Deviations:

________ Materials (test booklets, transparencies) were not marked.
Deviations:

________ Stopped within time schedule.
If not, note reason:

________ Additional comments on back.

Date ______________________________
Location ________________________
Leader ___________________________
# Participants __________________

III INSTRUCTION IN RELEVANCE JUDGMENT (11:05 - 11:15)

Start time ___________ Stop time ___________

_________ Started on time.
If not, note reason:

If participants entered late, note entry point in terms of page number on script.

JUDGES MAKE RELEVANCE JUDGMENT (11:15 - 11:50)

Start time ___________ Stop time ___________

_________ Started on time.
If not, note reason:

If participants entered late, note entry point in terms of page number on script.

_________ Leader followed script.
If not, note any deviations:

_________ Mechanical procedures went smoothly.
If problems, note the nature of the problem(s) and where they occured in the script.

_________ Participants' attention focussed on the task.
If not, note point(s) on the script where we "lose" people.

_________ Participants readily understood 4-point scale.
If not, note problems:

_________ Leader fielded questions easily.
Note all questions; identify by an (*) if there were problem questions by the leader.

_________ Stopped within time schedule.
If not, note reason:

_________ Additional comments on back.

Date ______________________
Location ___________________
Leader _____________________
# Participants _____________

IV INSTRUCTION IN STANDARD SETTING (11:55 - 12:05)

Start time __________ Stop time __________

________ Started on time.
If not, note reason:

If participant(s) entered late, note entry point in terms of page number on script:

FIRST STANDARD SETTING SESSION (12:05 - 12:40)

________ Started on time.
If not, note reason:

If participant(s) entered late, note entry point in terms of page number on script:

________ Leader followed script.
If not, note any deviations:

________ Mechanical procedures went smoothly.
If problems, note the nature of the problem(s) where they occured in the script.

________ Participants attention focused on the task.
If not, note point(s) on the script where we "lose" people:

________ Participants understood probability estimates.
If not, note problems:

________ Participants marked record booklets correctly.
If not, note problems:

________ Leader fielded questions easily.
Note all questions; identify by an (*) if there were problem questions for the leader.

________ Materials were not taken from the testing room, and materials were secured during lunch break.
If not, note problems:

________ Stopped within time schedule.
If not, note reason:

________Additional comments on back.

Date ____________________________
Location ________________________
Leader __________________________
# Participants _______________

V MINI-TEST EXERCISE (2:00 - 2:25)

Start time __________ Stop time __________

________ Started on time.
If not, note reason:
If participant(s) entered late, note entry point in terms of page number on script.

________ Leader followed script.
If not, note any deviations:

________ Mechanical procedures went smoothly.
If problems, note the nature of the problem(s) and where they occurred in the script.

________ Participants attention focussed on the task.
If not, note points on the script where we "lose" people:

________ Similation exercise was successful.
If not, note problem(s):

_____ NTE Exercise--% answered correctly (specify problems):

_____ NTE Exercise--30 - 50% correct (specify problems):

_____ NTE Exercise - Standard (specify problems):

________ Leader fielded questions easily.
Note all questions; identify by an (*) any that were problem questions for the leader.

________ Stopped within time schedule.
If not, note reason:

________ Additional comments on back.

Date ______________________________
Location _________________________
Leader ___________________________
No. Participants ________________

Participants ________________

VI SECOND STANDARD-SETTING (2:25 - 3:10)

Start time ____________ Stop time ____________

______ Started on time.
If not, note reason:
If participant(s) entered late, note entry point in terms of page number on script.

______ Leader followed script.
If not, note any deviations:

______ Mechanical procedures went smoothly.
If problems, note the nature of the problem(s) and where they occurred in the script.

______ Participants attention focussed on the task.
If not, note point(s) on the script where we "lose" people:

______ Leader fielded questions easily.
Note all questions; identify by an (*) any that were problem questions for the leader:

______ All materials collected from all participants.
If not, note deviations:

______ Stopped within time schedule.
If not, note reason:

______ Additional comments on back.

Date ____________________
Location ____________________
Leader ____________________
# Participants ____________________

VII PUBLIC SCHOOL FACULTY DEBRIEFING SESSION (3:10 - 3:25)

Start time __________ Stop time __________

________Started on time.
If not, note reason:
If participant(s) entered late, note entry point in terms of page number on script:

________ Leader followed script.
If not, note any deviations:

________ Participants attention focussed on the task.
If not, note point(s) on the script where we "lose" people:

________ Leader fielded questions easily.
Note all questions; identify by an (*) any that were problem questions for the leader.

________ Stopped within time schedule.
If not, note reason:

________ Additional comments on back.

Date ____________________
Location ________________
Leader __________________
# Participants __________

VIII INSTRUCTION IN INSTRUCTIONAL VALIDITY (3:25 - 3:30)

Start time ________ Stop time ________

________ Started on time.
If not, note reason:
If participant(s) entered late, note entry point in terms of page number on script:

INSTRUCTIONAL VALIDITY JUDGMENTS (3:30 - 4:10)

Start time ________ Stop time ________

________ Started on time.
If not, note reason:
If participant(s) entered late, note entry point in terms of page number on script.

________ Leader followed script.
If not, note any deviations:

________ Mechanical procedures went smoothly.
If problems, note the nature of the problem(s) and where they occurred in the script.

________ Participants attention focussed on the task.
If not, note point(s) on the script where we "lose" people:

________ Leader fielded questions easily.
Note all questions; identify by an (*) any that were problem questions for the leader:

________ Participants readily understood Validity questions.
If not, note problems:

________ Stopped within time schedule.
If not, note reason:

________ Additional comments on back.

Date ____________________
Location ________________
Leader __________________
# Participants ___________

IX FINAL DEBRIEFING (4:10 - 4:25)

Start time __________ Stop time __________

________ Started on time.
If not, note reason:
If participant(s) entered late, note entry point in terms of page number on script:

________ Leader followed script.
If not, note any deviations:

________ Participants attention focussed on the task.
If not, note point(s) on the script where we "lose" people:

________ Leader fielded questions easily.
Note all questions; identify by an (*) any that were problem questions for the leader.

________ All payment forms correctly filled out.
If not, note problems.

________ Stopped within the time schedule.
If not, note reason:

________ Additional comments on back.

Behavioral Observation Procedures. In every data-collection session, group assistants were instructed to observe and to record adherence to and deviations from planned procedures (a copy of these forms is included in the Materials section, pages 174-183 of this document). Resulting data were carefully analyzed, to determine the fidelity of data-collection activities to planned procedures. the results of the data. Overall, the group assistants reported that scripts were followed closely by group leaders, that procedures were followed as established, and that data-collection sessions progressed smoothly. Some problems were encountered at each site, but they were rectified and at most, caused the sessions to run longer than planned. The most serious problems encountered are listed below, for each data-collection site.

A. Asheville

1. Most sessions started late because participants arrived late.

2. Reading Test I was improperly assembled by ETS. The ETS representative corrected the error, and had new tests mailed for the Graham and Goldsboro sessions.

3. A participant in the Literature/Fine Arts session left early due to a personal matter.

4. Some materials were misassembled in the Listening, Science, and Writing multiple-choice sessions. These problems were corrected, with a slight time delay.

B. Graham

1. Sessions started late because of late-arriving participants.

2. Lunch was delayed and caused some sessions to get behind in the afternoon.

C. Goldsboro

1. Seven participants were about 45 minutes late. These individuals were able to catch up with their groups by working during the first break and with special attention from group assistants.

2. Several of the test areas contained two or three testing groups in one large room. As the sessions began,some participants and leaders complained about having difficulty hearing. As the sessions proceeded, the groups learned to accomodate to these conditions, and procedures went smoothly.

Data reduction procedures

The sources of data provided by participants in the three data collection sessions in October 1983 were of three basic types. Forms with written comments by the judges are discussed in another section of this report. The other two types of data forms were computer-readable forms and hand-input forms. The processing of both computer-readable and hand-input forms is discussed in this section.

All the data were read into a Teletek micro-computer. In order to prepare the forms for this process, the following procedure was implemented:

a) The names on all data forms were checked individually for appropriate social security numbers (matching with the number as indicated on a master roster sheet). The computer program used social security numbers to collate responses by individual judges.

b) All computer-readable forms were hand-checked for extraneous marks, incomplete erasures and double markings of items (these ambiguous replys were erased and treated as "non- responses").

c) The forms which were computer-read on-site at the time of data collection were given the same close attention as listed in a) and b) above.

d) The forms were separated into six groups, according to a planned input sequence:

1. Standard setting forms, Session 1 [computer-readable]- one set for each multiple-choice subtest.
2. Standard setting forms, Session 2 [computer-readable]-

one set for each multiple-choice subtest.

3. -Item relevance/instructional validity forms [computer-readable]

   -Global evaluation of instructional validity forms [college/university faculty only; hand-input]

   - Test blueprint forms [college/university faculty only; hand-input]

   A set of these forms was assembled for each multiple-choice subtest.

4. Evaluation response forms [computer-readable]

5. -Essay, 1st judgment session forms,Set 1 [computer-readable]

   -2nd judgment session form, Set 1 [hand-input]

   -3rd judgment session form, Set 1 [hand-input]

   -Content/relevance form [hand-input]

6. -Essay, 1st judgment session form, Set 2 [computer-readable]

   -2nd judgment session form, Set 2 [hand-input]

   -3rd judgment session form, Set 2 [hand-input]

The forms in Groups 3, 5 and 6 above were reorganized by putting the forms for individual participants together. The computer program was designed to accept the data from each of these forms sequentially in a single computer run. For each group of forms, each participant was identified as a teacher at a public school or a college/university faculty member and each institution was identified with a code number.

Entering the data into the computer was a reasonably smooth process. BASIC computer programs used to input data from the computer-readable forms and hand-input forms were written by Dr. Robert Muir of the UNC-G physics department. The programs were extensively tested and debugged using small data samples before being used to process the entire batch of forms. The programs were written to allow re-entering of hand-input data and re-reading of computer-readable forms when there were inappropriate marks. The few errors which were made in typing in the data were carefully noted, and later corrected using the Wordstar word processing program.

Some of the data files [standard setting data, groups 1, 2, 5 and 6 above] were analysed using BASIC computer programs run on the Teletek micro-computer. Other data files [item relevance/ instructional validity and evaluation data, groups 3 and 4 above] were transfered using a communication link program via a modem to a Vax 11/780 [UNC-G Academic Computer Center] so as to analyze the data using the SPSS-X computer package.

Listings of computer programs used to read, reduce and edit data on test standards, test validity and participants' evaluations of procedures follow. Each of these Computer programs was written by Dr. Robert Muir.

```
10 REM*************************************
20 REM           STDFILE.BAS
25 REM
26 REM  WRITES DISKETTE FILES OF DATA FROM STANDARD FORMS
30 REM
40 REM   WRITTEN BY:  BOB MUIR
50 REM     VERSION OF 20 DEC '83 (4:30PM)
100 REM ***********INITIALIZATION*************
101 DIM WT(11),RESP(55),HIST(45,11),DA(64)
105 L2=LOG(2)     :REM USED IN SUBROUTINE 10000
110 UNLOCKR$=CHR$(17)
120 EOB$="!"
130 BEL$=CHR$(7)+CHR$(7)+CHR$(7)
135 CLR$=CHR$(12)
140 LINEF$=CHR$(10)
145 FEED$=CHR$(12)
150 FOR I=1 TO 7
160 : READ TNAME$(I)
170 : READ TLEN(I)
180 NEXT I
190 DATA "READING",30,"WRITING",45,"LISTENING",40,"MATH",25,
    "SCIENCE",30
200 DATA "SOC. STUDIES",30,"LIT./F.A.",35
210 FOR I=0 TO 10
220 : READ WT(I)
230 NEXT I
240 DATA 0,.05,.155,.255,.355,.455,.555,.655,.755,.855,.955
250 H$="**************************************************"
260 REM
261 REM    SET UP BUBBLE DECODING ARRAY
262 REM
263 FOR I=0 TO 63
264 :DA(I)=9
265 NEXT I
266 DA(0)=0:DA(1)=1:DA(2)=2:DA(4)=3:DA(8)=4:DA(16)=5:DA(32)=6
300 PRINT CLR$  :  REM CLEAR SCREEN
340 PRINT LINEF$;LINEF$;LINEF$
350 PRINT TAB(10);"1  READING"
351 PRINT TAB(10);"2  WRITING"
352 PRINT TAB(10);"3  LISTENING"
353 PRINT TAB(10);"4  MATH"
354 PRINT TAB(10);"5  SCIENCE"
355 PRINT TAB(10);"6  SOCIAL STUDIES"
356 PRINT TAB(10);"7  LITERATURE/FINE ARTS"
360 PRINT TAB(10);"8  QUIT"
368 PRINT LINEF$
370 INPUT"     ENTER NUMBER OF TEST TO BE PROCESSED ";TESTN
375 IF TESTN=8 THEN END
380 LENGTH=TLEN(TESTN)
400 PRINT LINEF$
410 INPUT"ENTER FILE NAME FOR OUTPUT (WITHOUT EXTENSION)";FILE$
411 FILEN$=FILE$+".NUM":FILEC$=FILE$+".CHR"
412 PRINT LINEF$
413 INPUT"OUTPUT DISK DRIVE: A OR B";DR$
415 FILEN$=LEFT$(DR$,1)+":"+FILEN$
```

```
416 FILEC$=LEFT$(DR$,1)+":"+FILEC$
420 PRINT LINEF$;"THE DATA FOR ";TNAME$(TESTN);" WILL BE WRITTEN
    TO FILES ";FILEN$;" & ";FILEC$
450 OPEN"O",#1,FILEN$
455 OPEN"O",#2,FILEC$
460 PRINT LINEF$
470 INPUT"WHAT IS THE SESSION NUMBER";SESSION
500 JDGC=1      :REM 1ST CHAR IN JUDGE FIELD
510 TSTC=19   :REM 1ST CHAR IN TEST #
520 BEGC=21   :REM 1ST CHAR IN PROBABILITY FIELD
530 HRDF=1
1000 REM
1001 REM************MAIN ROUTINE*****************
1002 REM
1010 FCNTR=0   :REM FORMS COUNTER
1011 GOTO 1012
1012 PRINT BEL$;LF$;LF$;BEL$;"ROCK THE SCANTRON POWER SWITCH TO
     THE MIDDLE POSITION"
1013 PRINT "AND THEN BACK TO THE UPPER POSITION";LF$
1014 INPUT "PRESS RETURN WHEN READY TO CONTINUE";R$
1015 INPUT "TYPE P IF PUBLIC SCHOOL OR F IF COLLEGE/UNIV FACULTY";
     TJUDG$
1016 IF TJUDG$="P" THEN TJUDG=1
1017 IF TJUDG$="F" THEN TJUDG=2
1020 PRINT UNLOCKR$    :REM UNLOCK FORM READER
1030 FCNTR=FCNTR+1
1040 INPUT FORM$
1050 IF FORM$=EOB$ THEN 5010
1060 LENF=LEN(FORM$)
1070 IF LENF<>110 THEN 4010   :REM BAD FORM LENGTH
1074 REM
1075 REM    GET JUDGE'S NUMBER
1076 REM
1080 JUDGE$=MID$(FORM$,JDGC,18)
1090 SS$=""
1100 ESS=1    :REM SOC SEC # ERROR FLAG
1110 FOR I=1 TO 18 STEP 2
1120 : D$=MID$(JUDGE$,I,1)
1130 : IF D$="@" THEN 1180
1140 : GOSUB 10000
1150 : IF D>6 THEN ESS=2
1155 D=D-1  :REM  BUBBLES RANGE FROM 0 TO 9
1160 : SS$=SS$+RIGHT$(STR$(D),1)
1170 : GOTO 1220
1180 : D$=MID$(JUDGE$,I+1,1)
1185 IF D$="@" THEN ESS=2
1190 : GOSUB 10000
1200 : IF D>4 THEN ESS=2
1205 D=D+5
1210 : SS$=SS$+RIGHT$(STR$(D),1)
1220 NEXT I
1230 SSN=VAL(SS$)
1240 ON ESS GOTO 1250,4110
1244 REM
1245 REM      GET TEST NUMBER
```

```
1246 REM
1250 TEST$=MID$(FORM$,TSTC,2)
1260 ETST=1     :REM TEST # ERROR FLAG
1270 D$=LEFT$(TEST$,1)
1280 IF D$="@" THEN 1330
1290 GOSUB 10000
1300 IF D<1 OR D>6 THEN ETST=2
1305 IF TESTN<>D THEN ETST=3
1320 GOTO 1380
1330 D$=RIGHT$(TEST$,1)
1340 IF D$="@" THEN ETST=2
1350 GOSUB 10000
1360 IF D>1 THEN ETST=2
1365 D=D+6
1370 IF TESTN<>D THEN ETST=3
1380 ON ETST GOTO 1410,4210,4250
1399 REM
1400 REM**********CHECK FIELDS AFTER RESPONSE FIELD*********
1401 REM
1410 ENDC=BEGC+2*LENGTH-2
1420 EEOF=1    :REM END OF FIELD ERROR FLAG
1430 FOR I=ENDC+2 TO 110
1440 : IF MID$(FORM$,I,1)<>"@" THEN EEOF=2
1450 NEXT I
1460 ON EEOF GOTO 1470,4310   :REM BAD END OF FORM
1469 REM
1470 REM***********DECODE RESPONSES******************
1471 REM
1480 FOR I=1 TO LENGTH
1490 : RESP(I)=0
1500 NEXT I
1510 VALRES=0
1520 INVAL=0
1525 NIL=0
1530 RESP$=MID$(FORM$,BEGC,2*LENGTH)
1535 J=1
1540 FOR I=1 TO LENGTH
1550 : CHR1$=MID$(RESP$,J,1)
1560 : CHR2$=MID$(RESP$,J+1,1)
1570 : IF CHR1$+CHR2$="@@" THEN 1750
1580 : IF CHR1$<>"@" THEN IF CHR2$<>"@" THEN 1720
1590 : IF CHR1$="@" THEN 1660
1600 : D$=CHR1$
1610 : GOSUB 10000
1620 : IF D<=0 OR D>=7 THEN 1720
1630 : RESP(I)=D
1640 : VALRES=VALRES+1
1650 : GOTO  1780
1660 : D$=CHR2$
1670 : GOSUB 10000
1680 : IF D<=0 OR D>=5 THEN 1720
1690 : RESP(I)=D+6
1700 : VALRES=VALRES+1
1710 : GOTO 1780
1720 : RESP(I)=11     :REM MULT MARKS OR NOT DECODABLE
```

```
1730 : INVAL=INVAL+1
1740 : GOTO 1780
1750 : VALRES=VALRES+1
1760 : NILRES(NIL)=I
1780 : J=J+2
1785 : HIST(I,RESP(I))=HIST(I,RESP(I))+1
1790 NEXT I
1799 REM
1800 REM**********QUALITY CHECK******************
1801 REM
1810 IF VALRES=LENGTH THEN 1900   :REM ALL OK.  COMPUTE STANDARD
1811 IF NULFLG=2 THEN 1819
1812 IF NIL = 0 THEN 1820
1813 PRINT BEL$;LINEF$;"JUDGE--";SS$;"  HAS";NIL;" NULL RESPONSES.
     CHECK ITEMS:";
1814 FOR K=1 TO NIL: PRINT NILRES(K): NEXT K
1815 PRINT"THEN REREAD FORM."
1816 NULFLG=2
1817 FCNTR=FCNTR-1
1818 GOTO 1011:  REM GO READ ANOTHER FORM
1819 NULFLG=1
1820 IF INVAL=LENGTH-VALRES THEN 3000   :REM SOME BAD RESPONSES
1830 PRINT BEL$;LINEF$;"PROBLEM AT STMS 1820.  VALRES=";VALRES;"
     INVAL=";INVAL
1840 PRINT TAB(24);"LENGTH=";LENGTH
1850 STOP
1899 REM
1900 REM**********COMPUTE STANDARD***************
1901 REM
1910 STANDARD=0
1920 NANSWER=0
1930 NIL=0
1940 FOR I=1 TO LENGTH
1950 : STANDARD=STANDARD+WT(RESP(I))
1960 : IF RESP(I)<>0 THEN 1990
1970 : NIL=NIL+1
1980 : GOTO 2000
1990 : NANSWER=NANSWER+1
2000 NEXT I
2005 STD$=STR$(STANDARD)
2010 IF NIL=0 THEN 2030
2012 IF NANSWER/LENGTH>=.75 GOTO 2020
2015 STANDARD=99.999
2018 GOTO 2030
2020 STANDARD=STANDARD*LENGTH/NANSWER
2025 STD$=STR$(STANDARD)+"    N"
2030 ON HRDF GOTO 2040,2060
2040 HRDF=2
2050 PRINT "  JUDGE";TAB(13);"STANDARD  LENGTH  NANSWER   ";
     TNAME$(TESTN)
2052 LPRINT FEED$
2054 LPRINT "THE FOLLOWING DATA HAS BEEN WRITTEN TO FILE---"FILE$;
     LINEF$;LINEF$
2055 LPRINT"  JUDGE";TAB(13);"STANDARD  LENGTH  NANSWER   ";
     TNAME$(TESTN)
```

```
2057 STRD=INT(STANDARD+.500001)
2060 PRINT SS$;TAB(13);STANDARD;TAB(24);LENGTH;TAB(33);NANSWER
2070 LPRINT SS$;TAB(13);STANDARD;TAB(24);LENGTH;TAB(33);NANSWER
2075 PRINT#1,SS$;",";TESTN;",";
2076 PRINT#2,SS$;",";STR$(TESTN);",";
2080 FOR I=1 TO LENGTH
2081 : W=WT(RESP(I))
2082 : W$=STR$(W)
2083 : PRINT#1,W;",";
2084 : IF LEN(W$)=5 GOTO 2095
2085 : LW=LEN(W$):X=5-LW
2086 : FOR K=1 TO X
2087 :  W$=W$+"0"
2088 : NEXT K
2095 : PRINT#2,W$;",";
2100 NEXT I
2110 PRINT#1,STANDARD;",";SESSION;",";TJUDG
2120 STD$=STR$(STANDARD)
2130 LS=LEN(STD$)
2140 IF LS=7 GOTO 2200
2150 FOR X=1 TO 7-LS
2160 : STD$=STD$+"0"
2170 NEXT X
2200 PRINT#2,STD$;",";STR$(SESSION)
2990 GOTO 1011    :REM GO READ ANOTHER FORM
2999 REM
3000 REM*********DEAL WITH BAD RESPONSES**************
3001 REM
3010 PRINT BEL$;LINEF$;"JUDGE--";SS$;"  ";INVAL;" BAD RESPONSES
     TO ITEMS";
3020 FOR I=1 TO LENGTH
3030 : IF RESP(I)<11 THEN 3050
3040 : PRINT I;
3045 : HIST(I,RESP(I))=HIST(I,RESP(I))-1
3050 NEXT I
3060 PRINT LINEF$
3065 FCNTR=FCNTR-1
3070 PRINT "CHECK THESE ITEMS FOR MULTIPLE MARKS.  IF NOT
     CORRECTABLE,"
3071 PRINT "PUT FORM ASIDE, OTHERWISE REREAD IT."
3200 GOTO 1011   :REM GO READ ANOTHER FORM
3999 REM
4000 REM**********ERROR HANDLING***********
4001 REM
4005 REM  PROBLEM WITH FORM LENGTH
4010 PRINT BEL$;LINEF$;"FORM LENGTH ERROR--";LENF;"  FIELDS READ"
4020 PRINT LINEF$;"EITHER REREAD FORM"
4030 PRINT TAB(8);"CORRECT & REREAD FORM"
4040 PRINT TAB(8);"PRESS END OF BATCH ON READER"
4050 FCNTR=FCNTR-1
4060 GOTO 1011 :  REM GO READ ANOTHER FORM
4099 REM
4100 REM   BAD SOC. SEC. NO.
4101 REM
4110 PRINT BEL$;LINEF$;"BAD SOC. SEC. NO.-- CHECK FOR MULTIPLE
```

```
     MARKS OR BLANKS AND REREAD FORM"
4120 PRINT "IF REPEAT FAILURE, PUT FORM ASIDE AND GO TO NEXT FORM."
4130 FCNTR=FCNTR-1
4140 GOTO 1011 :  REM GO READ ANOTHER FORM
4199 REM
4200 REM   BAD TEST ;NO.
4201 REM
4210 PRINT BEL$;LINEF$;"BAD TEST NO.-- CHECK FOR ;MULTIPLE MARKS
     AND REREAD FORM"
4220 PRINT"IF REPEAT FAILURE, PUT FORM ASIDE AND  GO TO NEXT FORM."
4230 FCNTR=FCNTR-1
4240 GOTO 1011 :  REM GO READ ANOTHER FORM
4250 PRINT BEL$;LINEF$;"TEST NUMBER INCONSISTENT WITH EXPECTED
     NUMBER."
4251 PRINT "CHECK FORM, CORRECT NUMBER AND REREAD IT."
4252 GOTO 1011
4299 REM
4300 REM   MARKS IN ITEMS AFTER RESPONSE FIELD
4301 REM
4310 PRINT BEL$;LINEF$;"MARKS IN AREA AFTER VALID RESPONSE FIELD."
4320 INORES=INT((BEGC-1)/2+LENGTH)-10
4330 PRINT"THERE SHOULD BE NO MARKS AFTER ITEM";INORES
4340 PRINT"CLEAN UP FORM AND REREAD IT.  IF REPEATED FAILURES,"
4345 PRINT"PUT ASIDE AND GO TO NEXT FORM."
4350 FCNTR=FCNTR-1
4360 GOTO 1011 : REM GO READ ANOTHER FORM
4999 REM
5000 REM    END OF BATCH ROUTINE
5001 REM
5010 FCNTR=FCNTR-1
5020 LPRINT LINEF$;LINEF$;"A TOTAL OF";FCNTR;"  FORMS HAVE BEEN
     PROCESSED."
5030 PRINT LINEF$;LINEF$;"A TOTAL OF";FCNTR;"  FORMS HAVE BEEN
     PROCESSED."
5100 CLOSE#1
5110 CLOSE#2
6000 REM
6010 PRINT LINEF$;LINEF$;LINEF$;"PRESS RETURN TO GET THE MENU";
6020 INPUT R$
6030 GOTO 300
10000 REM*******************************************
10010 REM          SUBROUTINE FOR CONVERSION        *
10020 REM        OF CHARACTER TO BUBBLE NUMBER      *
10030 REM           INPUT IS D$.  OUTPUT IS D       *
10040 REM*******************************************
10050 M=ASC(D$)-64
10060 D=DA(M)
10100 RETURN
60000 STOP
```

```
10 REM****************************************
20 REM            INSTVAL.BAS
25 REM
26 REM  WRITES DISKETTE FILES OF DATA FROM INST'L VAL'TY FORMS
27 REM          AND FROM GLOBAL AND BLUEPRINT FORMS
30 REM
40 REM   WRITTEN BY:  BOB MUIR
50 REM                20 DEC '83 (3:40 PM)
51 REM   CORRECTED: 19 DEC '83 (4:50 PM)
100 REM ***********INITIALIZATION**************
101 DIM RELEV(45), VALID(45), BLUE$(12), BLUE(12), RESP(45),DA(64)
110 UNLOCKR$=CHR$(17)
120 EOB$="!"
130 BEL$=CHR$(7)+CHR$(7)+CHR$(7)
135 CLR$=CHR$(12)
140 LF$=CHR$(10)
145 FEED$=CHR$(12)
150 FOR I=1 TO 7
160 : READ TNAME$(I)
170 : READ TLEN(I)
180 NEXT I
190 DATA "READING",30,"WRITING",45,"LISTENING",40,"MATH",25,
    "SCIENCE",30
200 DATA "SOC. STUDIES",30,"LIT./F.A.",35
210 REM
211 REM    SET UP ARRAY FOR DECODING CHARACTERS
212 REM
220 FOR I=0 TO 63
230 : DA(I)=9
235 NEXT I
240 DA(0)=0:DA(1)=1:DA(2)=2:DA(4)=3:DA(8)=4:DA(16)=5:DA(32)=6
290 HRDF=0
300 PRINT CLR$  :  REM CLEAR SCREEN
310 PRINT TAB(10);"ITEM RELEVANCE--INSTRUCTIONAL VALIDITY"
340 PRINT LF$;LF$;LF$
350 PRINT TAB(10);"1  READING"
351 PRINT TAB(10);"2  WRITING"
352 PRINT TAB(10);"3  LISTENING"
353 PRINT TAB(10);"4  MATH"
354 PRINT TAB(10);"5  SCIENCE"
355 PRINT TAB(10);"6  SOCIAL STUDIES"
356 PRINT TAB(10);"7  LITERATURE/FINE ARTS"
360 PRINT TAB(10);"8  QUIT"
368 PRINT LF$
370 INPUT"     ENTER NUMBER OF TEST TO BE PROCESSED ";TESTN
375 IF TESTN=8 THEN END
380 LENGTH=TLEN(TESTN)
400 PRINT LINEF$
410 INPUT"ENTER FILE NAME FOR OUTPUT (WITHOUT EXTENSION)";FILE$
411 FILEN$=FILE$+".NUM":FILEC$=FILE$+".CHR"
412 PRINT LINEF$
413 INPUT"OUTPUT DISK DRIVE: A OR B";DR$
415 FILEN$=LEFT$(DR$,1)+":"+FILEN$
416 FILEC$=LEFT$(DR$,1)+":"+FILEC$
420 PRINT LINEF$;"THE DATA FOR ";TNAME$(TESTN);" WILL BE WRITTEN
```

```
    TO FILES ";FILEN$;" & ";FILEC$
450 OPEN"O",#1,FILEN$
455 OPEN"O",#2,FILEC$
460 PRINT LINEF$
500 JDGC=1      :REM 1ST CHAR IN JUDGE FIELD
510 TSTC=19    :REM 1ST CHAR IN TEST #
520 BEGC=21    :REM 1ST CHAR IN PROBABILITY FIELD
1000 REM
1001 REM************MAIN ROUTINE*****************
1002 PRINT LF$;LF$;"PUT FIRST FORM INTO READER.";LF$
1010 FCNTR=0    :REM FORMS COUNTER
1020 PRINT UNLOCKR$     :REM UNLOCK FORM READER
1030 FCNTR=FCNTR+1
1040 INPUT FORM$
1050 IF FORM$=EOB$ THEN PRINT BEL$;LF$;"YOU SHOULD NOT PRESS END
     OF BATCH SWITCH!";GOTO 340
1060 LENF=LEN(FORM$)
1070 IF LENF<>110 THEN 4010    :REM BAD FORM LENGTH
1074 REM
1075 REM    GET JUDGE'S NUMBER
1076 REM
1080 JUDGE$=MID$(FORM$,JDGC,18)
1090 SS$=""
1100 ESS=1     :REM SOC SEC # ERROR FLAG
1110 FOR I=1 TO 18 STEP 2
1120 : D$=MID$(JUDGE$,I,1)
1130 : IF D$="@" THEN 1180
1140 : GOSUB 10000
1150 : IF D>6 THEN ESS=2
1155 D=D-1  :REM  BUBBLES RANGE FROM 0 TO 9
1160 : SS$=SS$+RIGHT$(STR$(D),1)
1170 : GOTO 1220
1180 : D$=MID$(JUDGE$,I+1,1)
1185 IF D$="@" THEN ESS=2
1190 : GOSUB 10000
1200 : IF D>4 THEN ESS=2
1205 D=D+5
1210 : SS$=SS$+RIGHT$(STR$(D),1)
1220 NEXT I
1230 SSN=VAL(SS$)
1240 ON ESS GOTO 1250,4110
1244 REM
1245 REM      GET TEST NUMBER
1246 REM
1250 TEST$=MID$(FORM$,TSTC,2)
1260 ETST=1     :REM TEST # ERROR FLAG
1270 D$=LEFT$(TEST$,1)
1280 IF D$="@" THEN 1330
1290 GOSUB 10000
1300 IF D<1 OR D>6 THEN ETST=2
1305 IF TESTN<>D THEN ETST=3
1320 GOTO 1380
1330 D$=RIGHT$(TEST$,1)
1340 IF D$="@" THEN ETST=2
1350 GOSUB 10000
```

```
1360 IF D>1 THEN ETST=2
1365 D=D+6
1370 IF TESTN<>D THEN ETST=3
1380 ON ETST GOTO 1410,4210,4250
1399 REM
1400 REM**********CHECK FIELDS AFTER RESPONSE FIELD*********
1401 REM
1410 ENDC=BEGC+2*LENGTH-2
1420 EEOF=1     :REM END OF FIELD ERROR FLAG
1430 FOR I=ENDC+2 TO 110
1440 : IF MID$(FORM$,I,1)<>"@" THEN EEOF=2
1450 NEXT I
1460 ON EEOF GOTO 1470,4310   :REM BAD END OF FORM
1469 REM
1470 REM***********DECODE RESPONSES*******************
1471 REM
1480 FOR I=1 TO LENGTH
1485 : RESP(I)=0
1490 : RELEV(I)=0
1495 : VALID(I)=0
1500 NEXT I
1510 VALRES=0
1520 INVAL=0
1525 NIL=0
1530 RESP$=MID$(FORM$,BEGC,2*LENGTH)
1535 J=1
1540 FOR I=1 TO LENGTH
1550 : CHR1$=MID$(RESP$,J,1)
1560 : CHR2$=MID$(RESP$,J+1,1)
1570 : IF CHR1$+CHR2$="@@" THEN 1750
1590 : IF CHR1$="@" THEN D=0:GOTO 1630
1595 REM      DEAL WITH LEFT FIELD
1600 : D$=CHR1$
1610 : GOSUB 10000
1620 : IF D<=0 OR D>=6 THEN 1720
1630 : RELEV(I)=D
1650 REM     DEAL WITH RIGHT FIELD
1660 : D$=CHR2$
1665 : IF D$="@" THEN D=0 : GOTO 1700
1670 : GOSUB 10000
1680 : IF D<=2 OR D>=7 THEN 1720
1690 : VALID(I)=D-3
1700 : VALRES=VALRES+1
1710 : GOTO 1780
1720 : RESP(I)=11     :REM MULT MARKS OR NOT DECODABLE
1730 : INVAL=INVAL+1
1740 : GOTO 1780
1750 : VALRES=VALRES+1
1760 : NILRES(NIL)=I
1780 : J=J+2
1790 NEXT I
1799 REM
1800 REM**********QUALITY CHECK******************
1801 REM
1810 IF VALRES=LENGTH THEN 1900   :REM ALL OK.  COMPUTE STANDARD
```

```
1811 IF NULFLG=2 THEN 1819
1812 IF NIL = 0 THEN 1820
1813 PRINT BEL$;LINEF$;"JUDGE--";SS$;"  HAS";NIL;" NULL RESPONSES.
     CHECK ITEMS:";
1814 FOR K=1 TO NIL: PRINT NILRES(K): NEXT K
1815 PRINT"THEN REREAD FORM."
1816 NULFLG=2
1817 FCNTR=FCNTR-1
1818 GOTO 1020:  REM GO READ ANOTHER FORM
1819 NULFLG=1
1820 IF INVAL=LENGTH-VALRES THEN 3000   :REM SOME BAD RESPONSES
1830 PRINT BEL$;LINEF$;"PROBLEM AT STMS 1820.  VALRES=";VALRES;"
     INVAL=";INVAL
1840 PRINT TAB(24);"LENGTH=";LENGTH
1850 STOP
1899 REM
1900 REM************* GET KEYBOARD INFO ******************
1901 REM
1910 PRINT BEL$;LF$;BEL$;"READY FOR KEYBOARD ENTRY.  FIRST ROCK
     THE SCAN-TRON POWER SWITCH"
1920 PRINT "    TO THE MIDDLE POSITION AND BACK TO THE UPPER
     POSITION.";LF$
1940 INPUT "WHAT IS THE SCHOOL CODE ";SCHOOL
1941 INPUT"IS THIS CORRECT (Y OR N)";R$
1942 IF LEFT$(R$,1)="N" GOTO 1940
1950 PRINT ""
1960 INPUT "WHAT IS THE DATE (4, 6, OR 8) ";DATE
1961 INPUT"IS THIS CORRECT (Y OR )N";R$
1962 IF LEFT$(R$,1)="N" GOTO 1960
1970 PRINT ""
1980 INPUT "IS JUDGE EDUCATION, DISCIPLINE, OR PUBLIC (E, D, OR P) ";
     JTYPE$
1981 INPUT"IS THIS CORRECT (Y OR N)";R$
1982 IF LEFT$(R$,1)="N" GOTO 1980
1983 IF JTYPE$="E" THEN JTYPE=1:GOTO 1990
1985 IF JTYPE$="D" THEN JTYPE=2:GOTO 1990
1987 JTYPE=3
1990 PRINT ""
2000 INPUT "ENTER SOCIAL SECURITY NO. ";SS1$
2010 IF SS1$=SS$ GOTO 2100
2020 PRINT BEL$;LF$;"THIS DOES NOT MATCH NO. READ FROM FORM. "
2021 PRINT "    PLEASE RE-ENTER SOC. SEC. NO. ";
2030 INPUT SS2$
2040 IF SS2$=SS$ GOTO 2100
2050 PRINT LF$;"THIS DOES NOT MATCH EITHER.  PLEASE CHECK ROSTER."
2052 PRINT "  THE NUMBER READ FROM THE FORM IS: ";SS$
2060 INPUT "IS THIS CORRECT (Y OR N)";R$
2070 IF LEFT$(R$,1)="Y" GOTO 2100
2080 PRINT LF$;"ENTER CORRECT SOC. SEC. NO. ";
2090 INPUT SS1$
2091 PRINT LF$;"ARE YOU SURE ABOUT THIS ONE (Y OR N) ";
2092 INPUT R$
2093 IF LEFT$(R$,1) ="N" GOTO 2080
2100 PRINT LF$;"ENTER BLUEPRINT DATA (TYPE B FOR BLANK OR E TO
     END IF LESS THAN 12 ITEMS)"
```

```
2105 FOR M=1 TO 12:BLUE$(M)=" ":BLUE(M)=0:NEXT M
2110 FOR M=1 TO 12
2120 : PRINT M;"  ";
2130 : INPUT BLUE$(M)
2140 : B$=LEFT$(BLUE$(M),1)
2142 : IF B$="E" THEN BLUE$(M)=" ":GOTO 2160
2145 : IF B$="Y" THEN BLUE(M)=1:GOTO 2150
2146 : IF B$="B" THEN BLUE(M)=0:BLUE$(M)=" ": GOTO 2150
2147 : BLUE(M)=2
2150 NEXT M
2160 PRINT LF$;"THE BLUEPRINT DATA IS:  ";
2165 PRINT "CHARACTERS: ";
2170 FOR M=1 TO 12
2180 : PRINT BLUE$(M);",";
2190 NEXT M
2191 PRINT CHR$(13);LF$;TAB(25);"NUMBERS:     ";
2192 FOR M=1 TO 12
2193 : PRINT BLUE(M);",";
2194 NEXT M
2200 PRINT LF$
2230 INPUT "IS THIS CORRECT (Y OR N)";R$
2240 IF LEFT$(R$,1) = "N" GOTO 2100
2250 PRINT LF$
2260 INPUT "ENTER GLOBAL INSTRUCTIONAL VALIDITY ";GLIV$
2270 INPUT "IS THIS CORRECT (Y OR N)";R$
2280 IF LEFT$(R$,1) = "N" GOTO 2250
2290 T$=LEFT$(GLIV$,1)
2292 IF T$="A" THEN GLIV=1:GOTO 2300
2293 IF T$="B" THEN GLIV=2:GOTO 2300
2294 IF T$="C" THEN GLIV=3:GOTO 2300
2295 GLIV=4
2300 REM
2310 REM            WRITE DATA TO FILES
2320 REM
2330 PRINT#1,SS$;",";TESTN;",";
2340 PRINT#2,SS$;",";STR$(TESTN);",";
2350 FOR I=1 TO 45
2360 : PRINT#1,RELEV(I);",";
2370 : PRINT#2,RIGHT$(STR$(RELEV(I)),1);",";
2380 NEXT I
2390 FOR I=1 TO 45
2400 : PRINT#1,VALID(I);",";
2410 : PRINT#2,RIGHT$(STR$(VALID(I)),1);",";
2420 NEXT I
2440 PRINT#1,DATE;",";JTYPE;",";
2450 PRINT#2,RIGHT$(STR$(DATE),1);",";JTYPE$;",";
2460 FOR M=1 TO 12
2470 : PRINT#1,BLUE(M);",";
2480 : PRINT#2,BLUE$(M);",";
2490 NEXT M
2500 PRINT#1,GLIV;",";SCHOOL
2510 PRINT#2,GLIV$;",";RIGHT$(STR$(SCHOOL),2)
2800 REM
2801 PRINT BEL$;LF$
2810 INPUT "READ ANOTHER FORM (Y OR N) ";R$
```

```
2820 IF LEFT$(R$,1)="Y" GOTO 1020
2830 PRINT LF$;"A TOTAL OF";FCNTR;" FORMS HAVE BEEN PROCESSED AND "
2831 PRINT "  WRITTEN TO FILES:  ";FILEN$;" &"
2840 PRINT TAB(21);FILEC$
2850 ON HRDF GOTO 2860,2880
2860 HRDF=2
2870 LPRINT FEED$;TAB(10);"ITEM RELEVANCE -- INSTRUCTIONAL VALIDITY"
2880 LPRINT LF$;"TEST";TESTN;":  ";FCNTR;" RECORDS WRITTEN TO
     FILES:  ";FILEN$;" & ";FILEC$
2890 CLOSE#1
2900 CLOSE#2
2910 PRINT LF$;LF$;"PRESS RETURN TO GET TO MENU."
2920 INPUT R$
2930 GOTO 300
2999 REM
3000 REM*********DEAL WITH BAD RESPONSES**************
3001 REM
3010 PRINT BEL$;LINEF$;"JUDGE--";SS$;"  ";INVAL;" BAD RESPONSES
     TO ITEMS";
3020 FOR I=1 TO LENGTH
3030 : IF RESP(I)<11 THEN 3050
3040 : PRINT I;
3050 NEXT I
3060 PRINT LINEF$
3065 FCNTR=FCNTR-1
3070 PRINT "CHECK THESE ITEMS FOR MULTIPLE MARKS.  IF NOT
     CORRECTABLE,"
3071 PRINT "PUT FORM ASIDE, OTHERWISE REREAD IT."
3200 GOTO 1020   :REM GO READ ANOTHER FORM
3999 REM
4000 REM**********ERROR HANDLING***********
4001 REM
4005 REM  PROBLEM WITH FORM LENGTH
4010 PRINT BEL$;LINEF$;"FORM LENGTH ERROR--";LENF;"  FIELDS READ"
4020 PRINT LINEF$;"EITHER REREAD FORM"
4030 PRINT TAB(8);"CORRECT & REREAD FORM"
4040 PRINT TAB(8);"PRESS END OF BATCH ON READER"
4050 FCNTR=FCNTR-1
4060 GOTO 1020 :  REM GO READ ANOTHER FORM
4099 REM
4100 REM   BAD SOC. SEC. NO.
4101 REM
4110 PRINT BEL$;LINEF$;"BAD SOC. SEC. NO.-- CHECK FOR MULTIPLE
     MARKS OR BLANKS AND REREAD FORM"
4120 PRINT "IF REPEAT FAILURE, PUT FORM ASIDE AND GO TO NEXT FORM."
4130 FCNTR=FCNTR-1
4140 GOTO 1020 :  REM GO READ ANOTHER FORM
4199 REM
4200 REM   BAD TEST ;NO.
4201 REM
4210 PRINT BEL$;LINEF$;"BAD TEST NO.-- CHECK FOR ;MULTIPLE MARKS
     AND REREAD FORM"
4220 PRINT"IF REPEAT FAILURE, PUT FORM ASIDE AND  GO TO NEXT FORM."
4230 FCNTR=FCNTR-1
4240 GOTO 1020 :  REM GO READ ANOTHER FORM
```

```
4250 PRINT BEL$;LINEF$;"TEST NUMBER INCONSISTENT WITH EXPECTED
     NUMBER."
4251 PRINT "CHECK FORM, CORRECT NUMBER AND REREAD IT."
4252 GOTO 1020
4299 REM
4300 REM   MARKS IN ITEMS AFTER RESPONSE FIELD
4301 REM
4310 PRINT BEL$;LINEF$;"MARKS IN AREA AFTER VALID RESPONSE FIELD."
4320 INORES=INT((BEGC-1)/2+LENGTH)-10
4330 PRINT"THERE SHOULD BE NO MARKS AFTER ITEM";INORES
4340 PRINT"CLEAN UP FORM AND REREAD IT.  IF REPEATED FAILURES,"
4345 PRINT"PUT ASIDE AND GO TO NEXT FORM."
4350 FCNTR=FCNTR-1
4360 GOTO 1020 : REM GO READ ANOTHER FORM
4999 REM
5000 REM    END OF BATCH ROUTINE
5001 REM
5010 FCNTR=FCNTR-1
5020 LPRINT LINEF$;LINEF$;"A TOTAL OF";FCNTR;"  FORMS HAVE BEEN
     PROCESSED."
5030 PRINT LINEF$;LINEF$;"A TOTAL OF";FCNTR;"  FORMS HAVE BEEN
     PROCESSED."
5100 CLOSE#1
5110 CLOSE#2
6000 REM
6010 PRINT LINEF$;LINEF$;LINEF$;"PRESS RETURN TO GET THE MENU";
6020 INPUT R$
6030 GOTO 300
10000 REM*******************************************
10010 REM          SUBROUTINE FOR CONVERSION        *
10020 REM        OF CHARACTER TO BUBBLE NUMBER      *
10030 REM           INPUT IS D$.  OUTPUT IS D       *
10040 REM*******************************************
10050 M=ASC(D$)-64
10060 D=DA(M)
10100 RETURN
60000 STOP
```

```
10 REM ************************************************************
20 REM     READING DATA FROM EVALUATION FORMS
30 REM
40 REM     WRITTEN BY:  BOB MUIR
50 REM                  5 JAN '84
60 REM                   3:09 PM
70 REM
80 REM
90 REM*************************************************************
100 REM                 INITIALIZATION
110 DIM VALR$(6),RESP(6),NBD(25),NUMDAT(25)
120 UNLOCKR$=CHR$(17)     :REM UNLOCK READER CHAR
130 EOB$="!"              :REM END OF BLOCK CHAR
140 BEL$=CHR$(7):BEL$=BEL$+BEL$:BEL$=BEL$+BEL$
150 CLR$=CHR$(12)         :REM CLEAR SCREEN CHAR
160 FEED$=CLR$            :REM FORM FEED CHAR
170 LF$=CHR$(10)          :REM LINE FEED CHAR
199 REM
200 FOR K=1 TO 6
210 :READ VALR$(K)        :REM SET UP VALID RESPONSES
220 :RESP(K)=K-1
230 NEXT K
240 DATA "@","A","B","D","H","P"
250 REM "@"=NO MARK, "A"=BUBBLE A, "B"=BUBBLE B, "D"= BUBBLE C
260 REM "H"=BUBBLE D, "P"=BUBBLE E
299 REM
300 TITLE$="PARTICIPANT'S EVALUATION"
310 PRINT CLR$;LF$;TAB(30);TITLE$;LF$
320 PRINT LF$
330 INPUT "HAS READER BEEN SET UP FOR 6 COLUMN FORMS (Y OR N)";R$
340 IF LEFT$(R$,1)="Y" GOTO 365
350 PRINT LF$;"THEN TURN IT OFF AND SET THE INTERFACE CONTROL
    SWITCH FOR 1-6."
360 END
365 PRINT CLR$;TAB(20);TITLE$;LF$;LF$
366 LPRINT FEED$;LF$;TAB(30);TITLE$
367 PRINT TAB(10);"1  READING"
368 PRINT TAB(10);"2  WRITING"
369 PRINT TAB(10);"3  LISTENING"
370 PRINT TAB(10);"4  MATH"
371 PRINT TAB(10);"5  SCIENCE"
372 PRINT TAB(10);"6  SOCIAL STUDIES"
373 PRINT TAB(10);"7  LITERATURE/FINE ARTS"
374 PRINT TAB(10);"8  ESSAY"
375 PRINT TAB(10);"9  QUIT";LF$;LF$
377 INPUT "ENTER NUMBER OF SUBJECT TO BE PROCESSED  ";SUBJ$
380 IF SUBJ$="9" THEN END
381 SUBJ=VAL(SUBJ$)
385 FILE$="B:EVAL"+SUBJ$+".NUM"
397 PRINT LF$;"DATA WILL BE WRITTEN TO FILE --- ";FILE$
398  LPRINT LF$;"DATA WILL BE WRITTEN TO FILE --- ";FILE$
400 OPEN "O",#1,FILE$
405 FCTR=0
410 PRINT LF$;LF$;"WHAT IS THE SOCIAL SECURITY NO. ON NEXT FORM
```

```
    OR TYPE END IF DONE";
420 INPUT SS$
425 IF LEFT$(SS$,1) ="E" GOTO 5000
430 PRINT CLR$;LF$;"IS THIS CORRECT (Y OR N)   ";LEFT$(SS$,3);"-";
    MID$(SS$,4,2);"-";RIGHT$(SS$,4);
440 INPUT R$
450 IF LEFT$(R$,1)="Y" GOTO 471
460 INPUT "RE-ENTER SOC. SEC. NO.";SS$
470 GOTO 430
471 PRINT
472 INPUT "TYPE P IF PUBLIC SCHOOL OR F IF COLLEGE/UNIV FACULTY";
    TJUDG$
473 IF TJUDG$="P" THEN TJUDG = 1
474 IF TJUDG$="F" THEN TJUDG = 2
480 PRINT LF$;"PUT FORM INTO READER AND PRESS RETURN ON KEYBOARD."
490 INPUT R$
500 PRINT UNLOCKR$                :REM UNLOCK READER
510 INPUT FORM$                   :REM READ FORM
520 IF FORM$=EOB$ GOTO 5000       :REM GO WRAP UP
530 FCTR=FCTR+1
540 LENF=LEN(FORM$)
550 IF LENF<>26 GOTO 4000         :REM BAD FORM LENGTH
1000 REM
1010 REM      DECODE RESPONSEX
1020 REM
1030 REM      CHECK FOR BAD RESPONSES
1040 NBAD=0
1050 FORM$=MID$(FORM$,2,25)
1060 FOR I=1 TO 25
1070 : FOR K=1 TO 6
1080 :    IF MID$(FORM$,I,1)=VALR$(K) GOTO 1110
1090 : NEXT K
1100 :NBAD=NBAD+1:NBD(NBAD)=I
1110 NEXT I
1120 IF NBAD=0 GOTO 1300
1130 REM    REPORT PROBLEMS WITH THIS FORM
1140 PRINT BEL$;LF$;"THERE ARE";NBAD;" INVALID RESPONSES.  CHECK
     ITEMS: ";
1150 FOR K=1 TO NBAD: PRINT NBD(K);: NEXT K: PRINT""
1160 FCTR=FCTR-1
1170 PRINT LF$;"EITHER:  1  CORRECT FORM AND REREAD IT OR"
1180 PRINT TAB(10);"2  PUT IT ASIDE AND GO TO THE NEXT FORM.";LF$
1190 INPUT"CHOICE (1 OR 2)";C
1200 ON C GOTO 420,1210
1210 LPRINT "DATA FOR ";SS$;" NOT WRITTEN TO DISK BECAUSE OF
     INVALID RESPONSES."
1220 GOTO 350
1300 REM
1310 REM          ANALYZE GOOD FORM
1320 REM
1330 FOR I=1 TO 25
9340 : FOR K=1 TO 6
1350 :    IF MID$(FORM$,I,1)<>VALR$(K) GOTO 1370
1360 :    NUMDAT(I)=K-1
1365 :    GOTO 1380
```

```
1370 : NEXT K
1380 NEXT I
1400 REM
1410 REM WRITE DATA TO DISK
1420 REM
1430 PRINT#1,SS$;SUBJ;",";TJUDG;
1440 FOR I=1 TO 25
1450 : PRINT#1,",";NUMDAT(I);
1460 NEXT I
1470 PRINT#1,""
1480 PRINT BEL$;LF$;LF$;BEL$;"ROCK THE SCAN-TRON POWER SWITCH TO
     THE MIDDLE POSITION AND THEN BACK TO"
1490 PRINT "THE UPPER POSITION.";LF$
1500 INPUT "PRESS RETURN WHEN READY TO CONTINUE.";R$
1510 GOTO 410
4000 REM
4010 REM                BAD FORM LENGTH
4020 REM
4030 PRINT BEL$;LF$;LF$;BEL$;"TIMING MARK PROBLEM.  THERE SHOULD
     BE 26, BUT";LENF;" WERE DETECTED."
4040 PRINT "EITHER REREAD FORM OR"
4050 PRINT TAB(8);"CORRECT & REREAD FORM OR"
4060 PRINT TAB(8);"ROCK SCAN-TRON POWER SWITCH TO THE MIDDLE
     POSITION, TYPE CONTROL-C AND GET HELP"
4070 FCTR=FCTR-1
4080 GOTO 420:  REM REREAD SAME FORM
5000 REM
5010 REM                WRAP UP PROCESSING
5020 REM
5030 CLOSE#1
5040 PRINT LF$;FCTR;" FORMS HAVE BEEN WRITTEN TO FILE---";FILE$
5050 LPRINT LF$;LF$;FCTR;" FORMS HAVE BEEN WRITTEN  TO FILE--";FILE$
5990 GOTO 365
```

```
10 REM*************************************************
20 REM      READING DATA FROM JUDGES OF ESSAY TEST
30 REM
40 REM      WRITTEN BY:  BOB MUIR
50 REM                   17 JAN '84
55 REM                   10:00 AM
60 REM
70 REM*************************************************
100 REM*****************INITIALIZATION******************
110 DIM RESP(11,6),NBD(11)
115 DIM NUMDAT(11),CHRDAT$(11),ACC2(11),ACC2$(11),UAC2(11),UAC2$(11)
116 DIM ACC3(11),ACC3$(11),UAC3(11),UAC3$(11)
120 UNLOCKR$=CHR$(17):REM UNLOCK READER CHAR.
130 EOB$="!"          :REM END OF BLOCK CHAR.
140 BEL$=CHR$(7):BEL$=BEL$+BEL$:BEL$=BEL$+BEL$
150 CLR$=CHR$(12)     :REM CLEAR SCREEN CHAR
155 FEED$=CHR$(12)    :REM FORM FEED CHAR
160 LF$=CHR$(10):LF$=LF$+LF$:REM DOUBLE LINE FEED
199 REM
200 REM INITIALIZE FOR ANSWERS TO BLOCKS 4 & 5 QUESTIONS
210 FOR I=1 TO 6
220 :READ Q1$(I)
230 NEXT I
240 DATA "SA","A","U","D","SD",""
290 FOR K=1 TO 6:REM SET UP VALID RESPONSE CHARS
310 :READ VALR$(K)
320 NEXT K
330 DATA "@","A","B","D","H","P"
340 REM "@"=NO MARK, "A"=BUBBLE A MARKED, "B"=B MARKED
342 REM "D"=C MARKED, "H"=D MARKED, "P"=E MARKED
400 REM
410 TITLE$="COMMUNICATIONS SKILLS EVALUATION"
420 PRINT CLR$;LF$;TAB(20);TITLE$
430 LPRINT FEED$;TAB(25);TITLE$
470 PRINT LF$;"ENTER NUMBER CORRESPONDING TO ONE OF THE FOLLOWING."
472 PRINT LF$;TAB(10);"1  PROCESS ESSAY SET #1"
473 PRINT TAB(10);"2  PROCESS ESSAY SET #2"
475 PRINT TAB(10);"3  QUIT"
477 PRINT LF$;"CHOICE ";
480 INPUT SETN
490 IF SETN=3 THEN END
495 SETN$=RIGHT$(STR$(SETN),1)
500 PRINT " "
510 NFILE$="B:ESAYSET"+SETN$+".NUM"
520 CFILE$="B:ESAYSET"+SETN$+".CHR"
530 PRINT LF$;"DATA FOR ESSAY SET";SETN;" WILL BE WRITTEN TO FILES:  ";
    NFILE$
531 PRINT TAB(50);CFILE$
533 LPRINT LF$;"DATA FOR ESSAY SET";SETN;" WILL BE WRITTEN TO FILES:  ";
    NFILE$
534 LPRINT TAB(50);CFILE$
540 PRINT LF$;LF$;"IS THE CORRECT DISKETTE IN DRIVE B";
550 INPUT Z$
555 IF LEFT$(Z$,1)="Y" GOTO 560
557 PRINT LF$;"PUT THE DISKETTE IN DRIVE B AND PRESS RETURN WHEN
```

```
    READY TO CONTINUE."
558 INPUT Z$
560 OPEN"O",#1,NFILE$
570 OPEN"O",#2,CFILE$
574 REM
575 REM PRINT HEADER RECORD TO BOTH FILES
576 REM
580 PRINT#1,"ESSAY SET,";SETN;","
590 PRINT#2,"ESSAY SET,";SETN$;","
1000 REM*************************************************
1001 REM                MAIN ROUTINE
1002 REM*************************************************
1800 FCTR=0:REM  FORMS COUNTER
1900 PRINT LF$;LF$
1910 INPUT "ENTER SOC. SEC. NO. (OR END)  ";SS$
1915 IF LEFT$(SS$,1)="E" GOTO 7110
1920 PRINT CLR$
1930 PRINT "IS THIS CORRECT   ";LEFT$(SS$,3);"-";MID$(SS$,4,2);"-";
     RIGHT$(SS$,4);"    (Y OR N)";
1940 INPUT Z$
1950 IF LEFT$(Z$,1)="Y" GOTO 1960
1953 INPUT "RE-ENTER SOC. SEC. NO.";SS$
1955 GOTO 1920
1960 PRINT LF$;LF$;"WHAT IS THE SESSION DATE (4, 6 OR 8)";
1970 INPUT DATE$
1980 DATE=VAL(DATE$)
1983 PRINT LF$
1985 INPUT "WHAT IS THE SCHOOL ID NO.";SCHOOL
1987 SCHOOL$=STR$(SCHOOL)
1988 IF LEN(SCHOOL$)=3 THEN SCHOOL$=RIGHT$(SCHOOL$,2)
1990 INPUT "IS THIS CORRECT";Z$
1993 IF LEFT$(Z$,1)<>"Y" GOTO 1985
2010 PRINT LF$;LF$;BEL$;"PUT FORM INTO READER."
2020 PRINT UNLOCKR$:REM UNLOCK FORM READER
2030 INPUT FORM$
2040 IF FORM$=EOB$ GOTO 7110:REM ALL DONE
2050 FCTR=FCTR+1
2060 LENF=LEN(FORM$)
2070 IF LENF<>26 GOTO 7020:REM BAD FORM LENGTH
2080 REM
2090 REM             DECODE RESPONSES
2100 REM
2110 REM CHECK FOR BAD RESPONSES
2120 NBAD=0
2130 FORM$=MID$(FORM$,2,11)
2140 FOR I=1 TO 11
2145 :GOOD=0
2150 :FOR K=1 TO 6
2160 : IF MID$(FORM$,I,1)=VALR$(K) THEN GOOD =1
2170 :NEXT K
2175 :IF GOOD=1 GOTO 2190
2180 :NBAD=NBAD+1:NBD(NBAD)=I
2190 NEXT I
2200 IF NBAD=0 GOTO 2270
2210 REM
```

```
2220 REM     REPORT PROBLEMS WITH THIS FORM
2230 REM
2240 PRINT BEL$;LF$;"THERE ARE";NBAD;" IMPROPER RESPONSES.  CHECK
     ITEMS";
2250 FOR K=1 TO NBAD:PRINT NBD(K);:NEXT K:PRINT
2260 GOTO 7050
2270 REM
2280 REM     ANALYZE GOOD FORM
2290 REM
2300 FOR I=1 TO 11
2310 :FOR K=1 TO 6
2320 : IF MID$(FORM$,I,1)<>VALR$(K) GOTO 2330
2322 : CHRDAT$(I)=RIGHT$(STR$(K),1)
2324 : NUMDAT(I)=K
2330 :NEXT K
2340 NEXT I
2350 PRINT LF$;BEL$;"ROCK THE SCAN-TRON POWER SWITCH TO THE MIDDLE
     POSITION AND THEN BACK TO"
2352 PRINT "THE UPPER POSITION.";LF$
2353 INPUT "PRESS RETURN WHEN READY TO CONTINUE.";Z$
2529 REM
2530 REM GET VALUES FROM KEYBOARD FOR BLOCK 2
2531 REM
2532 PRINT LF$;LF$;"INPUT DATA FROM THE STANDARD SETTING SESSION 2
     FORM.";LF$
2533 INPUT "ENTER SOC. SEC. NO.";SS1$
2534 IF SS1$=SS$ GOTO 2540
2535 PRINT "SOC. SEC. NO. DOESN'T MATCH THE ONE FOR THE MARK SENSE
     FORM."
2536 INPUT "RE-ENTER SOC. SEC. NO.";SS1$
2537 GOTO 2534
2540 PRINT LF$;"ENTER ETS SCORES JUDGED ACCEPTABLE.  ENTER 99 FOR
     FIRST BLANK."
2545 SKIP=0
2550 FOR I=1 TO 11
2555 :IF SKIP=1 THEN ACC2(I)=99:GOTO 2571
2560 :PRINT I;
2570 :INPUT "  ";ACC2(I)
2571 :ACC2$(I)=STR$(ACC2(I))
2572 :IF LEN(ACC2$(I))=3 THEN ACC2$(I)=RIGHT$(ACC2$(I),2)
2575 :IF ACC2(I)=99 THEN SKIP=1
2580 NEXT I
2584 PRINT LF$;"ESSAYS JUDGED ACCEPTABLE:";
2585 FOR I=1 TO 11
2586 :PRINT ACC2(I);
2587 NEXT I
2588 PRINT LF$;"ARE THESE CORRECT (Y OR N)";
2589 INPUT Z$
2590 IF LEFT$(Z$,1)<>"Y" GOTO 2540
2591 PRINT LF$;LF$;"WHAT IS THE ACCEPTABLE ESSAY WITH THE LOWEST
     SCORE";
2592 INPUT LOW2
2600 INPUT "IS THIS CORRECT (Y OR N)";Z$
2610 IF LEFT$(Z$,1)<>"Y" GOTO 2591
2620 LOW2$=STR$(LOW2)
```

```
2630 IF LEN(LOW2$)=3 THEN LOW2$=RIGHT$(LOW2$,2)
2770 PRINT LF$;"ENTER ETS SCORES JUDGED UNACCEPTABLE.  ENTER 99
     FOR BLANK."
2772 SKIP=0
2780 FOR I=1 TO 11
2785 :IF SKIP=1 THEN UAC2(I)=99:GOTO 2802
2790 :PRINT I;
2800 :INPUT "  ";UAC2(I)
2802 :UAC2$(I)=STR$(UAC2(I))
2803 :IF LEN(UAC2$(I))=3 THEN UAC2$(I)=RIGHT$(UAC2$(I),2)
2805 :IF UAC2(I)=99 THEN SKIP=1
2810 NEXT I
2814 PRINT LF$;"ESSAYS JUDGED UNACCEPTABLE:";
2815 FOR I=1 TO 11
2816 :PRINT UAC2(I);
2817 NEXT I
2818 PRINT LF$;"ARE THESE CORRECT (Y OR N)";
2819 INPUT Z$
2820 IF LEFT$(Z$,1)<>"Y" GOTO 2770
2821 PRINT LF$;LF$;"WHAT IS THE UNACCEPTABLE ESSAY WITH THE HIGHEST
     SCORE";
2822 INPUT HIGH2
2830 INPUT "IS THIS CORRECT (Y OR N)";Z$
2853 IF LEFT$(Z$,1)<>"Y" GOTO 2821
2860 HIGH2$=STR$(HIGH2)
2870 IF LEN(HIGH2$)=3 THEN HIGH2$=RIGHT$(HIGH2$,2)
2900 REM GET VALUES FROM KEYBOARD FOR BLOCK 3
2910 PRINT LF$;LF$;"INPUT DATA FROM THE STANDARD SETTING SESSION
     3 FORM.";LF$
2920 INPUT "ENTER SOC. SEC. NO.";SS1$
2930 IF SS1$=SS$ GOTO 2970
2940 PRINT "SOC. SEC. NO. DOESN'T MATCH THE ONE FOR THE MARK SENSE
     FORM."
2950 INPUT "RE-ENTER SOC. SEC. NO.";SS1$
2960 GOTO 2930
2970 PRINT LF$;"ENTER ETS SCORES JUDGED ACCEPTABLE.  ENTER 99 FOR
     FIRST BLANK."
2975 SKIP=0
2980 FOR I=1 TO 11
2985 :IF SKIP=1 THEN ACC3(I)=99:GOTO 3003
2990 :PRINT I;
3000 :INPUT "  ";ACC3(I)
3003 :ACC3$(I)=STR$(ACC3(I))
3004 :IF LEN(ACC3$(I))=3 THEN ACC3$(I)=RIGHT$(ACC3$(I),2)
3010 :IF ACC3(I)=99 THEN SKIP=1
3020 NEXT I
3060 PRINT LF$;"ESSAYS JUDGED ACCEPTABLE:";
3070 FOR I=1 TO 11
3080 :PRINT ACC3(I);
3090 NEXT I
3100 PRINT LF$;"ARE THESE CORRECT (Y OR N)";
3110 INPUT Z$
3120 IF LEFT$(Z$,1)<>"Y" GOTO 2970
3130 PRINT LF$;LF$;"WHAT IS THE ACCEPTABLE ESSAY WITH THE LOWEST
     SCORE";
```

```
3140 INPUT LOW3
3150 INPUT "IS THIS CORRECT (Y OR N)";Z$
3200 IF LEFT$(Z$,1)<>"Y" GOTO 3130
3203 LOW3$=STR$(LOW3)
3205 IF LEN(LOW3$)=3 THEN LOW3$=RIGHT$(LOW3$,2)
3210 PRINT LF$;"ENTER ETS SCORES JUDGED UNACCEPTABLE.  ENTER 99
     FOR FIRST BLANK."
3215 SKIP=0
3220 FOR I=1 TO 11
3225 :IF SKIP=1 THEN UAC3(I)=99:GOTO 3243
3230 :PRINT I;
3240 :INPUT UAC3(I)
3243 :UAC3$(I)=STR$(UAC3(I))
3245 :IF LEN(UAC3$(I))=3 THEN UAC3$(I)=RIGHT$(UAC3$(I),2)
3250 :IF UAC3(I)=99 THEN SKIP=1
3260 NEXT I
3300 PRINT LF$;"ESSAYS JUDGED UNACCEPTABLE:";
3310 FOR I=1 TO 11
3320 :PRINT UAC3(I);
3330 NEXT I
3340 PRINT LF$;"ARE THESE CORRECT (Y OR N)";
3350 INPUT Z$
3360 IF LEFT$(Z$,1)<>"Y" GOTO 3210
3370 PRINT LF$;LF$;"WHAT IS THE UNACCEPTABLE ESSAY WITH THE HIGHEST
     SCORE";
3380 INPUT HIGH3
3390 INPUT "IS THIS CORRECT (Y OR N)";Z$
3430 IF LEFT$(Z$,1)<>"Y" GOTO 3370
3440 HIGH3$=STR$(HIGH3)
3450 IF LEN(HIGH3$)=3 THEN HIGH3$=RIGHT$(HIGH3$,2)
3500 REM
3501 REM GET KEYBOARD DATA FOR BLOCK 4 (SKIP BLOCKS 4 & 5 IF
     ESSAY SET 2)
3502 REM
3510 IF SETN=2 GOTO 5000
3520 PRINT LF$;LF$;"ENTER SOC. SEC. NO. FROM CONTENT/RELEVANCY FORM:";
3522 PRINT "IF NO FORM, JUST PRESS RETURN."
3530 INPUT SS1$
3533 IF SS1$<>SS$ GOTO 3540
3534 B4Q1$="6"
3535 B4Q1=6
3536 B4Q2$="5"
3537 B4Q2=5
3540 IF SS1$=SS$ GOTO 3580
3550 PRINT "SOC. SEC. NO. DOESN'T MATCH THE ONE FOR THE MARK SENSE
     FORM."
3560 INPUT "RE-ENTER SOC. SEC. NO.";SS1$
3570 GOTO 3540
3580 PRINT LF$;"ENTER RESPONSE TO QUESTION 1.";
3590 INPUT B4Q1$
3595 B4Q1=0
3600 FOR I=1 TO 6
3610 :IF B4Q1$=Q1$(I) THEN B4Q1=I
3620 NEXT I
3621 IF B4Q1<>0 GOTO 3635
```

```
3623 PRINT BEL$;LF$;BEL$;"YOUR RESPONSE--";B4Q1$;" IS INVALID."
3625 GOTO 3580
3635 B4Q1$=STR$(B4Q1)
3640 INPUT "IS THIS CORRECT (Y OR N)";Z$
3650 IF LEFT$(Z$,1)<>"Y" GOTO 3580
3660 PRINT LF$;"ENTER RESPONSE TO QUESTION 2.   1  CRUCIAL"
3670 PRINT TAB(33);"2  IMPORTANT"
3680 PRINT TAB(33);"3  QUESTIONABLE"
3690 PRINT TAB(33);"4  IRRELEVANT"
3700 PRINT TAB(33);"5  NO RESPONSE"
3710 INPUT "WHICH";B4Q2
3720 INPUT "IS THIS CORRECT (Y OR N)";Z$
3730 IF LEFT$(Z$,1)<>"Y" GOTO 3660
3735 B4Q2$=STR$(B4Q2)
3740 REM
3750 REM GET KEYBOARD DATA FOR BLOCK 5
3760 REM
3770 PRINT LF$;LF$;"ENTER SOC. SEC. NO. FROM INSTRUCTIONAL VALIDITY
     FORM."
3772 PRINT "IF NO FORM, JUST PRESS RETURN."
3780 INPUT SS1$
3783 IF SS1$<>"" GOTO 3790
3784 B5Q1$="6"
3785 B5Q1=6
3786 B5Q2$="6"
3787 B5Q2=6
3788 GOTO 5000
3790 IF SS1$=SS$ GOTO 3830
3800 PRINT "SOC. SEC. NO. DOESN'T MATCH THE ONE FOR THE MARK
     SENSE FORM."
3810 INPUT "RE-ENTER SOC. SEC. NO.";SS1$
3820 GOTO 3790
3830 PRINT LF$;LF$;"ENTER RESPONSE TO QUESTION 1";
3840 INPUT B5Q1$
3850 INPUT "IS THIS CORRECT (Y OR N)";Z$
3860 IF LEFT$(Z$,1)<>"Y" GOTO 3830
3865 B5Q1=0
3870 FOR I=1 TO 6
3880 :IF B5Q1$=Q1$(I) THEN B5Q1=I
3890 NEXT I
3891 IF B5Q1<>0 GOTO 3903
3893 PRINT BEL$;LF$;BEL$;"YOUR RESPONSE--";B5Q1$;" IS INVALID."
3895 GOTO 3830
3903 B5Q1$=STR$(B5Q1)
3910 PRINT LF$;LF$;"ENTER RESPONSE TO QUESTION 2";
3920 INPUT B5Q2$
3930 INPUT "IS THIS CORRECT (Y OR N)";Z$
3940 IF LEFT$(Z$,1)<>"Y" GOTO 3910
3945 B5Q2=0
3950 FOR I=1 TO 6
3960 :IF B5Q2$=Q1$(I) THEN B5Q2=I
3970 NEXT I
3971 IF B5Q2<>0 GOTO 3985
3973 PRINT BEL$;LF$;BEL$;"YOUR RESPONSE--";B5Q2$;" IS INVALID."
3975 GOTO 3910
```

```
3985 B5Q2$=STR$(B5Q2)
5000 REM
5010 REM OUTPUT BLOCK 1 TO DISK FILES
5020 REM
5030 PRINT#1,SS$;",";DATE;",";SCHOOL;",";
5040 PRINT#2,SS$;",";DATE$;",";SCHOOL$;",";
5050 PRINT SS$;",";DATE$;",";SCHOOL$;",";
5060 FOR I=1 TO 11
5070 :PRINT#1,NUMDAT(I);",";
5080 :PRINT#2,CHRDAT$(I);",";
5090 :PRINT CHRDAT$(I);",";
5100 NEXT I
5102 REM
5103 REM OUTPUT BLOCK 2
5104 REM
5105 PRINT#1,2;",";
5106 PRINT#2,"2";",";
5107 PRINT "2";",";
5110 FOR I=1 TO 11
5120 :PRINT#1,ACC2(I);",";
5150 :PRINT#2,ACC2$(I);",";
5160 :PRINT ACC2$(I);",";
5170 NEXT I
5180 PRINT#1,LOW2;",";
5220 PRINT#2,LOW2$;",";
5230 PRINT LOW2$;",";
5240 FOR I=1 TO 11
5250 :PRINT#1,UAC2(I);",";
5290 :PRINT#2,UAC2$(I);",";
5300 :PRINT UAC2$(I);",";
5310 NEXT I
5350 PRINT#1,HIGH2;",";
5360 PRINT#2,HIGH2$;",";
5370 PRINT HIGH2$;",";
5400 REM
5410 REM OUTPUT BLOCK 3
5420 REM
5430 PRINT#1,3;",";
5440 PRINT#2,"3";",";
5445 PRINT 3;",";
5450 FOR I=1 TO 11
5460 :PRINT#1,ACC3(I);",";
5470 :PRINT#2,ACC3$(I);",";
5475 :PRINT ACC3$(I);",";
5480 NEXT I
5490 PRINT#1,LOW3;",";
5500 PRINT#2,LOW3$;",";
5505 PRINT LOW3$;",";
5510 FOR I=1 TO 11
5520 :PRINT#1,UAC3(I);",";
5530 :PRINT#2,UAC3$(I);",";
5535 :PRINT UAC3$(I);",";
5540 NEXT I
5550 PRINT#1,HIGH3;",";
5560 PRINT#2,HIGH3$;",";
```

```
5565 PRINT HIGH3$;",";
5600 REM
5610 REM  OUTPUT BLOCK 4
5620 REM
5630 IF SETN=2 GOTO 7000
5640 PRINT#1,B4Q1;",";B4Q2;",";
5650 PRINT#2,B4Q1$;",";B4Q2$;",";
5660 PRINT B4Q1$;",";B4Q2$;",";
5700 REM
5710 REM OUTPUT BLOCK 5
5720 REM
5730 PRINT#1,B5Q1;",";B5Q2
5740 PRINT#2,B5Q1$;",";B5Q2$
5750 PRINT B5Q1$;",";B5Q2$
7000 GOTO 1900:REM GO READ ANOTHER FORM
7010 REM
7020 REM     BAD FORM LENGTH
7030 REM
7040 PRINT BEL$;LF$;"TIMING MARK PROBLEM.  THERE SHOULD BE 26,
     BUT";LENF;" WERE DETECTED."
7050 PRINT LF$;"EITHER  REREAD FORM OR CORRECT & REREAD THE FORM."
7072 PRINT LF$;"YOU MUST ROCK THE SCAN-TRON POWER SWITCH TO THE
     MIDDLE POSITION AND THEN "
7073 PRINT "BACK TO THE UPPER POSITION BEFORE PROCEEDING.
     PRESS RETURN WHEN READY."
7074 INPUT Z$
7080 FCTR=FCTR-1
7090 GOTO 1900:REM GO READ ANOTHER FORM
7100 REM
7110 REM    END OF BATCH.  REPORT RESULTS
7120 REM
7130 PRINT BEL$;LF$;LF$;"A TOTAL OF";FCTR;" SETS OF DATA HAVE BEEN
     PROCESSED FOR ESSAY SET ";SETN
7370 CLOSE#1
7380 CLOSE#2
7390 PRINT BEL$;LF$;LF$;"PROCESS ANOTHER SET";
7400 INPUT R$
7410 IF LEFT$(R$,1)="Y" GOTO 470
7420 END
```

Data analysis procedures

Analysis of validity data for multiple-choice and essay tests.

a. Relevance Judgements

Multiple Choice Tests: Each of the seven multiple- choice sub-tests in Reading, Writing, Listening, Science, Social Studies, Mathematics, Literature and Fine Arts were reviewed for content relevance in the same way. Participants rated each test item on a five-point scale. They were asked to assess the relevance of the skill or knowledge assessed by the item, to successful performance in the approved teacher education program with which they were most familiar. The four rating scale points included "Crucial", "Important", "Questionable Importance", and "Not Relevant". Participants could also indicate that they were unable to make the judgment. The panels consisted of college and university faculty and public school teachers.

The following descriptive statistics are reported for each test item:

a) The frequency distribution of the valid and missing responses,

b) The mean and standard deviation of the valid (non-missing) responses.

These statistics are presented in Tables 15 to 21 for each of the multiple choice tests.

In addition, descriptive statistics are reported for each of the tests taken as a whole:

a) The frequency distribution of the valid and missing responses, across all items on the test.

b) The mean of the valid (non-missing) responses for the total test.

These statistics are reported at the bottom of Tables15 through 21.

Essay Test: Writing skills were assessed by two sub-tests in the National Teacher Examinations Core Battery, a multiple-choice Writing subtest and an essay sub-test. Since there are no individual items for the Essay sub-test, an alternative method for assessing the relevance of the skills measured by the Essay sub-test to success in teacher education was developed. Two questions were asked of the participants. The first was a Likert-style question and it stated that "The NTE Essay Examination which I have reviewed, including its administration and scoring procedures, is a satisfactory measure of general writing skill." Participants were asked to mark a scale ranging from 1 = "strongly agree" to 5 = "strongly disagree" with 3 = "undecided". Results are reported in Table 22.

The second essay question asked the participants to consider the writing skills which were measured by the NTE Essay sub-test, and then to judge how important those skills were to success in the teacher education program with which they were most familiar. They were asked how important it would be for a student to have minimal skills in writing, as measured by the NTE Essay Test. They responded with the same choices available for relevance ratings on the multiple- choice tests ie. "crucial", "important", etc. Following this question participants were encouraged to write the reasons for their response.

The following descriptive statistics are reported in Table 22 for each of the two questions:

a) The frequency distribution of the valid and missing responses.

b) The mean and standard deviation for each question.

c) The percentages corresponding to valid and missing responses.

In addition, relevance judgements were analyzed for college and university faculty separately from those of public school judges. Results for the multiple choice sub-tests are given in Table 24, and for the Essay sub-test in Table 25.

b. Instructional Validity Judgements

Multiple Choice Tests: There are four different sources of information on the instructional validity of the multiple-choice tests: 1. individual item judgements, 2. global evaluation ratings, 3. blueprint rating data and 4) document analysis data. The nature of each source of data and the analysis procedures used, are described in the sections which follow. The college/university participants provided Instructional Validity judgements; public school teachers did not.

1. Individual Item Judgements

Participants were asked whether 90% or more of the students at their institution would have had a reasonable opportunity to secure instruction in the knowledge or skill assessed by each test item. They responded: "yes", "no" or "I can't make a judgement".

The following descriptive statistics are reported for each item on each of the seven multiple-choice tests.

a) The frequency distribution and percentages of the "yes" responses; and the frequency and percentages of the missing responses are reported in Tables 26 and 27.

b) The percentage of items on the test for which a majority of participants answered "yes" to the 90% question is reported in Table 31.

c) The average, standard deviation, and range of the percent of items marked "yes" by individual participants are reported in Table 28 and Table 31.

2. Global Evaluation Ratings

Each participant was asked to complete a single global evaluation question regarding the test he or she reviewed. The question asked the participants to assess the degree of similarity between the test and the relevant portions of the curriculum at their institution. They responded on a five-point scale ranging from 1 = "very similar" to 4 = "little similarity" and 5 = "blank".

The following descriptive statistics are reported for each test panel:

a) The distribution of frequencies and associated percentages for the valid and blank responses. These data are reported in Table 29.

b) The mean, standard deviation and range on the four- point scale are reported in Table 28.

3. Blueprint Rating Data

Each participant was presented with a copy of the test "blueprint" for the test which he or she were reviewing and was directed to answer the question, "Would 90% or more of the applicants to the teacher education programs at my institution have had a reasonable opportunity to secure instruction in the skills or knowledge defined by each statement in the blueprint prior to application for admission to teacher education?" The question was answered for each major topic on the blueprint either as "yes" or "no".

The following descriptive statistics were produced for each test and are reported in Table 29: the distribution of frequencies and associated percentages for each part of the test blueprint for each test.

4. A document analysis was performed during Phase II of the project and was described by Dr. D. Michelle Irwin, in Volume 3 of the Final Report dated June 30, 1983. The document analysis was designed to provide additional evidence from a slightly different perspective on the instructional validity of the multiple choice tests. Volume 3 should be consulted for a detailed description of the methodology and results of that analysis. Briefly, the procedure was designed to evaluate the opportunity to receive instruction afforded to students who take a "typical" sequence of courses in the teacher training institutions in North Carolina. Each of the institutions was requested to randomly sample a set of students, each of which had 48 to 75 semester hours completed at the institution. The set of individuals was assumed to be representative of the population of individuals applying for

admission to teacher education, and who would take the NTE Tests at the institution. Transcripts were collected for each individual and sent to the investigator for further analysis. A set of "typical" courses taken by students was identified and each course was associated with a particular NTE test. Judges then used materials describing the courses to determine the likelihood that the topic tested by the exam was taught in the course. Judgments ranged from 1, small likelihood, to 3, highly likely. In doing this, the test blueprint, which listed major topics on the test, was used to make that individual judgment for each major test topic. This information on instructional validity differed from that described above, in that judges different from those who delivered the instruction, provided the judgments, and secondly, that the judgments were based on courses actually taken by students in teacher education programs. There were several disadvantages to the procedure: first, judges in the document analysis were using limited print descriptions of courses for judgments and second, the rate of participation of different institutions in the process was not optimal. Only 25 institutions returned transcripts before the closing date for data analysis; some institutions would not participate because they were concerned about privacy violations. It should be pointed out that earlier cooperation in this phase of the project was quite good. Thirty-six institutions provided curriculum descriptions and advising materials. In any case, the representativeness of the evidence from the document analysis must be regarded with some caution.

Results of the Document Analysis were reported in Volume 3 as weighted averages on a 3 point scale for each dimension of the test blueprint. Both the central tendency (mean and median) and variability (range) of the weighted averages were examined. The evidence on the instructional validity for the total test was classified as "low"- less than 1.75, "moderate"- scores equal to 1.75 and less than or equal to 2.25, and "high"- greater than 2.25, using the three-point scale. This evidence is summarized and reported in Table 31.

All of the information describing the instructional validity for each of the seven multiple-choice tests is summarized in six indices and reported in Table 31. Table 31a provides a detailed definition of each of the indices.

Essay Sub-test: It was not possible to design an item-based series of instructional validity questions as was the case for the multiple-choice sub-tests. In order to assess the instructional validity of the Essay Test two Likert-type questions were asked. They attempted to determine whether students would have had a reasonable opportunity to receive instruction in writing skills both as described by ETS in its blueprint for the Essay sub-test (Question 1) or in the skills as actually assessed by the Essay sub-test (Question 2). Participants were provided with two statements and asked to indicate whether they "strongly agreed", "agreed", "disagreed","strongly disagreed" or were "uncertain".

The following descriptive statistics are reported in Table 30 with a copy of the questions asked.

a) The distribution of frequencies and associated percentages of responses to each question.

c.. Summarization of Validity Evidence

The basic approach taken to assessing the validity of the NTE Communication Skills and General Knowledge Tests was to assess multiple sources of evidence which would help to answer validity questions. We recognized that there was no single index which would serve to answer the validity question definitively. However, we felt that examination of multiple indices of validity, each perhaps somewhat fallible, would, in the cumulative sense, help to "triangulate" to a reasonable answer. For this reason multiple indices of validity were defined: five indices for the relevance of the multiple choice sub-tests and two for the relevance of the Essay sub-test; six indices of instructional validity for the multiple choice sub-tests and two for the instructional validity of the Essay sub-test. Those indices were arbitrary in the sense that there is nothing about any one of them that demands they be used to evaluate validity: however it must be emphasized that they are not capricious. To the contrary, they were the result of systematic study of previous research investigations of a similar nature, (for example, Lawrence Cross's work validating the NTE for the State of Virginia) or previous work which had been accepted by the courts, (for example, Educational Testing Service studies of the NTE for the State of North Carolina). These indices were carefully selected to provide relatively independent sources of information.

Relevance Indices: The following information is reported separately in Table 23, which describes the multiple -choice subtests and their relevance to success in teacher education. Each of the summative indices is operationally defined in detail in Table 23a.

a) The percentage of items on the test with mean ratings equal to or greater than 2.25.

b) The percentage of the ratings which were "not relevant".

c) The percentage of responses marked "crucial" or "important" out of the total number of valid responses for a test.

d) The percentage of items on the test which the majority of raters marked "crucial" or "important". That is, the number of items for which 51% or more of the valid ratings were marked "crucial" or "important", was divided by the number of test items.

e) The average of all responses on the 1-4 scale, reported in each of Tables 15 - 21.

f) The percentage of valid and missing responses to each of the tests, corresponding to the frequencies described in Tables 15 - 21. The percent of valid responses which are "crucial", "important", etc. is based on the total non-missing valid responses. The percent of missing responses is based on the total number of panelists asked to make relevance judgements.

In addition, the following information about the sub-test and panel is given:

a) The inter-rater reliability for each test.

b) The number of panelists and test length.

The Essay sub-test evidence was used to construct two indices, each of which were based on the two questions reported on in Table 22. The second question provides evidence on the relevance of the sub-test content to success in teacher education, and is relatively more important than question one. The first question provides evidence on what has traditionally been termed "content validity" in the psychometric literature. The essay test indicators consisted of the percentage of responses expressing an opinion in response to either question which were "strongly agree" or "agree" (Question 1) and the percentage of valid responses which were "crucial" or "important" (Question 2).

All of the indices varied considerably in their scale characteristics, therefore a decision model had to be designed which could be used to classify the data from the five indices as providing "satisfactory", "marginally satisfactory", or "unsatisfactory" evidence in support of the relevance of the sub-test. The criteria used to evaluate each of the revelance indices as "satisfactory", "unsatisfactory", or "marginally satisfactory" are reported in Table 23a. We tried to be as objective as possible in evaluating the validity evidence, as well as exercising fundamental common sense and good judgment in our decisions. The results of using those rules to evaluate each of the relevance indices is shown in the left hand portion of Table 32. A "+" is used to signify satisfactory evidence, a "?" to signify marginally satis-factory evidence, and a "-" to signify unsatisfactory evidence. Results are presented for the multiple-choice and the Essay sub-test. The table provides a convenient way to view a complex combination of results in totality.

Instructional Validity Indices: As was the case for the relevance evidence, the instructional validity of the sub-tests was evaluated using six indices of validity for the multiple-choice tests and two for the Essay sub-test. Operational definitions of each of the indices are given in Table 31a and each of the indices are reported in Table 31. The two Essay sub-test indices were based on the two Likert-type questions.

In order to evaluate each index in a systematic fashion, a decision model was established. The model provided clear rules by which the index would be classified as satisfactory, marginally satisfactory, or unsatisfactory. These are summarized in Table 31a.

The results of using these rules are reported in Table 32. Each of the indices is classified as "satisfactory" (+), "marginally satisfactory" (?) or as "unsatisfactory" (-), in support of the instructional validity of the Communication Skills and General Knowledge Tests.

Analysis of Standard-Setting Data

Overview. As described in Section II. C., above, recommendations on appropriate standards for all multiple-choice subtests of the NTE General Knowledge and Communication Skills Tests were collected in a common form. Recommendations on appropriate standards for the Essay subtest were collected through procedures that differed in detail, but were common in spirit, with those used for the multiple-choice subtests. Although recommendations on appropriate test standards were collected in several data-collection sessions, only those collected during the final sessions (the second session in the case of multiple-choice subtests and the third session in the case of the Essay subtest) were used as a basis for recommending standards on the NTE General Knowledge and Communication Skills Tests.

Analysis of participants` recommendations on appropriate standards involved several steps. Initial analyses were confined to individual subtests. Subsequent analyses combined data across subtests, to derive recommendations for the NTE General Knowledge and Communication Skills Tests. One set of analytic procedures was applied to data from the multiple-choice subtests. Another was applied to data from the Essay subtest.

The first step in analyzing data from the multiple-choice tests was to identify test items with mean relevance ratings of 2.25 or less. These items were eliminated from the computation of participants` recommended standards. The second step in the analysis was to compute a recommended standard on each subtest for each participant, by summing the partcipants` recommendations for each item in the subtest that had a mean relevance value of at least 2.25. The third step in the analysis involved computation of various descriptive statistics of the distribution of participants` recomended standards on each subtest. Statistics for individual subtests were combined across subtests by using a weighting formula supplied by Educational Testing Service. These weighted "raw" recommended standards were converted to ETS scale scores by applying the same conversion equations ETS uses to compute individual examinee`s scale scores. Finally, these statistics were used to construct a variety of tables and graphs.

Procedures used to analyze standards recommended for the Essay subtest were, except for the first few steps, quite similar to those used with the multiple-choice tests. Of course, recommendations provided for the Essay subtest did not involve any item-by-item data. Therefore, it was not necessary to screen items for relevance and then compute a recommendation for the entire subtest. The basic data used to derive a recommended standard on this subtest were participants` judgments on the ETS-assigned score of the lowest acceptable essay in each of two sets. These judgments were combined statistically, both across participants and sets of essays. The resulting statistics were then analyzed in conjunction with corresonding statistics resulting from analysis of recommendations on the multiple-choice subtests.

Analyses Used with Multiple-Choice Tests. Recall that each of the participants assigned to a multiple-choice subtest was asked to examine each item in that subtest and then make the following judgment:

"You are to imagine a hypothetical applicant who has the *minimum* amount of knowledge and skill in the subject area of the [ ] test, necessary to succeed in the North Carolina teacher education program *with which you are most familiar*. You might think of this hypothetical applicant as a 'low C' student, at least in the areas of skill and knowledge assessed by the [ ] section of the NTE.

"You are to estimate, for each item, the probability or 'chances' that this minimally competent applicant would be able to answer the test item correctly."

Participants indicated their judgments for each item by darkening a circle on an optically-scannable response sheet. Possible responses for each item ranged from "1" to "10". Participants were told that by marking "1" they would be estimating a probability of success on the item that was between zero and 0.10; by marking "2" they would be estimating a probability of success between 0.11 and 0.20; and so on. A mark of "10" represented a probability estimate between 0.91 and 1.00.

The computer program used to read data from the optically-scannable forms and create a magnetic disk file included a mathematical function for interpreting and translating marks within the "1" to "10" range. Each mark was converted to the midpoint of the range of probabilities it represented. Thus a mark of "1" was converted to a probability of 0.05; a mark of "2" was converted to a probability of 0.155; a mark of "3" was converted to a probability of 0.255; and so on, up to a mark of "10," which was converted to a probability of 0.955.

The first step in analyzing these data was to compute a recommended standard for each participant who provided judgments for a given subtest. Each participant`s recommended standard was computed by identifying all items in the subtest with mean relevance ratings of 2.25 or less, and then summing the estimated probabilities for all other items. This procedure was based on Angoff`s (1971) standard-setting method and had the added advantage that applicants to a teacher education program would not be held accountable for test items that were judged to be irrelevant.

A computer program was written in the BASIC compiler language to complete the item relevance screening, compute a recommended standard for each participant who judged items on each multiple-choice subtest, and create a floppy disk file of the resulting recommendations. A listing of this computer program, titled STANDESC.BAS, is shown below:

```
10 REM *****************************************************************
20 REM                         STANDESC.BAS
30 REM A PROGRAM TO COMPUTE DESCRIPTIVE STATISTICS ON MULTIPLE-CHOICE
   TEST
40 REM STANDARDS FOR NTE TESTS -- PHASE III OF THE QUALITY ASSURANCE
   PROGRAM
50DIM SS$(100),ZZ(100),X(100,45),STAN(100),P(100),Z(100),F(100),
   G(100),CUM(100),Y#(100)
60 DIM STDN(50),VARN(50),MEANN(50),SUMMX(50),SUMMSQ(50),D(50),NT(50)
61 DIM PUCO(50)
70 LPRINT
80 REM WRITTEN BY RICHARD M. JAEGER
90 PRINT
100 REM LAST REVISED AT 12:30 P.M. ON 6 JANUARY 1984
110 PRINT
120 PRINT "BE SURE THAT THE PRINTER IS CONNECTED AND ON!!!!"
130 REM THIS PROGRAM COMPUTES BASIC DESCRIPTIVE STATISTICS
140 REM IT COMPUTES A SAMPLE MEAN
150 REM IT COMPUTES A SAMPLE STANDARD DEVIATION
160 REM IT COMPUTES AN EXCLUSIVE RANGE
170 REM IT ASSUMES THAT DATA ARE PROVIDED IN THE FORM OF
180 REM AN UNORDERED ARRAY OF VALUES
190 REM IT LISTS THE VALUES OF THESE STATISTICS ON THE SCREEN
200 PRINT "INPUT A HEADING, NOT EXCEEDING 60 CHARACTERS"
```

```
210 INPUT H$
220 PRINT H$
230 PRINT "INPUT THE SAMPLE SIZE"
240 INPUT N
250 PRINT "THE SAMPLE SIZE IS";N
260 PRINT "INPUT THE SENSITIVITY OF MEASUREMENT"
270 PRINT "THAT IS, INPUT THE NUMBER OF PLACES TO THE RIGHT"
280 PRINT"             OF THE DECIMAL FOR THE DATA"
290 INPUT SENS
300 INPUT "ARE DATA TO BE READ FROM A FILE?";RESP$
310 PRINT "THE SENSITIVITY IS";SENS
320 IF LEFT$(RESP$,1)="Y" GOTO 1750
330 PRINT "INPUT THE DATA VALUES, ONE AT A TIME, IN RESPONSE"
340 PRINT "TO THE '?' AND HIT 'RETURN'"
350 FOR I=1 TO N
360 INPUT STAN(I)
370 NEXT I
380 SUMX=0
390 C=0
400 FOR J=1 TO NI
410 SUMMX(J)=0
420 NEXT J
430 SUMSQ=0
440 FOR J=1 TO NI
450 SUMMSQ(J)=0
460 NEXT J
470 FOR I=1 TO N
480 IF STAN(I)>55 GOTO 510
490 SUMX=SUMX+STAN(I)
500 SUMSQ = SUMSQ + (STAN(I)*STAN(I))
510 IF STAN(I)>55 THEN C=C+1
520 NEXT I
530 NS=N-C
540 MEAN=SUMX/NS
550 GOTO 760
560 FOR J = 1 TO NI
570 D(J) = 0
580 NEXT J
590 FOR I=1 TO N
600 FOR J = 1 TO NI
610 IF X(I,J)=0 GOTO 640
620 SUMMX(J)=SUMMX(J)+X(I,J)
630 SUMMSQ(J)=SUMMSQ(J)+(X(I,J)*X(I,J))
640 IF X(I,J)=0 THEN D(J) = D(J) + 1
650 NEXT J
660 NEXT I
670 FOR J=1 TO NI
680 NT(J) = N - D(J)
690 MEANN(J)=SUMMX(J)/NT(J)
700 STDN(J)=SUMMSQ(J) - (SUMMX(J)*SUMMX(J)/NT(J))
710 VARN(J)=STDN(J)/(NT(J)-1)
720 STDN(J)=SQR(STDN(J)/(NT(J)-1))
730 NEXT J
740 IF NBEL=0 GOTO 1970
750 GOTO 1870
```

```
760 STD# = SUMSQ - (SUMX*SUMX/NS)
770 VAR = STD#/(NS-1)
780 STD = SQR(STD#/(NS-1))
790 SEM = STD/SQR(NS)
800 VEM = SEM*SEM
810 PRINT H$
820 LPRINT H$
830 PRINT
840 LPRINT
850 PRINT "THE MEAN EQUALS"; MEAN
860 LPRINT "THE MEAN EQUALS";MEAN
870 FOR J=1 TO NI
880 PRINT "FOR ITEM NUMBER  ";J;"  THE MEAN RECOMM. STANDARD IS  ";
    MEANN(J)
890 LPRINT "FOR ITEM NUMBER  ";J;"  THE MEAN RECOMM. STANDARD IS  ";
    MEANN(J)
900 NEXT J
910 PRINT
920 PRINT "THE VARIANCE EQUALS"; VAR
930 LPRINT "THE VARIANCE EQUALS"; VAR
940 FOR J = 1 TO NI
950 PRINT "FOR ITEM NUMBER  ";J;" THE VARIANCE OF REC. STD. IS  ";
    VARN(J)
960 LPRINT "FOR ITEM NUMBER  ";J; " THE VARIANCE OF REC. STD. IS  ";
    VARN(J)
970 NEXT J
980 PRINT
990 PRINT "THE STANDARD DEVIATION EQUALS"; STD
1000 LPRINT "THE STANDARD DEVIATION EQUALS"; STD
1010 PRINT
1020 LPRINT
1030 PRINT "THE STANDARD ERROR OF THE MEAN EQUALS"; SEM
1040 LPRINT "THE STANDARD ERROR OF THE MEAN EQUALS"; SEM
1050 PRINT
1060 LPRINT
1070 LPRINT "THE VARIANCE ERROR OF THE MEAN EQUALS";VEM
1080 FOR J=1 TO NI
1090 PRINT "FOR ITEM NUMBER ";J;"  THE STD. DEV. OF REC. STD. IS";
     STDN(J)
1100 LPRINT "FOR ITEM NUMBER  ";J;"  THE STD. DEV. OF REC. STD. IS";
     STDN(J)
1110 NEXT J
1120 FOR I = 1 TO N
1130 Y#(I) = STAN(I)
1140 NEXT I
1150 FLIPS=1
1160 WHILE FLIPS
1170 FLIPS=0
1180 FOR I = 1 TO N-1
1190 IF Y#(I)>Y#(I+1) THEN SWAP Y#(I),Y#(I+1):FLIPS=1
1200 NEXT I
1210 WEND
1220 MIN= Y#(1)
1230 MAX = Y#(NS)
1240  RANGE = MAX - MIN
```

```
1250 PRINT
1260 PRINT "THE SMALLEST AND LARGEST VALUES ARE ";   MIN; "AND";
     MAX
1270 LPRINT "THE SMALLEST AND LARGEST VALUES ARE"; MIN; "AND"; MAX
1280 PRINT
1290 PRINT "THE EXCLUSIVE RANGE IS"; RANGE
1300 LPRINT "THE EXCLUSIVE RANGE IS"; RANGE
1310 PRINT
1320 INCL = RANGE + 2*(1/(2*10^SENS))
1330 PRINT "THE INCLUSIVE RANGE IS"; INCL
1340 LPRINT "THE INCLUSIVE RANGE IS"; INCL
1350 FOR I=1 TO N
1360 F(I)=1
1370 NEXT I
1380 FOR I=1 TO N
1390 FOR J=I+1 TO N
1400 IF Y#(J)=Y#(I) AND F(J)>0 THEN F(I)=F(I)+1
1410 IF Y#(J)=Y#(I) AND F(J)>0 THEN F(J)=F(J)-1
1420 NEXT J
1430 NEXT I
1440 FOR I=1 TO N
1450 IF F(I)=0 THEN Y#(I)=0
1460 NEXT I
1470 K=1
1480 FOR I=1 TO N
1490 IF Y#(I)>0 THEN Z(K)=Y#(I)
1500 IF Y#(I)>0 THEN G(K)=F(I)
1510 IF Y#(I)>0 THEN K=K+1
1520 NEXT I
1530 FOR L=1 TO K-1
1540 NEXT L
1550 CUM(1)=G(1)
1560 FOR L=2 TO K-1
1570 CUM(L) = CUM(L-1) + G(L)
1580 NEXT L
1585 Q25 = N/4
1587 Q75 = (3/4)*N
1590 Q=N/2
1600 YY=1
1602 WW=1
1604 VV=1
1610 FOR L=1 TO K-1
1620 IF CUM(L)<=Q THEN YY=YY+1
1622 IF CUM(L)<=Q25 THEN WW=WW+1
1624 IF CUM(L)<=Q75 THEN VV = VV+1
1630 NEXT L
1640 LRL = Z(YY) - (.5*1/10^SENS)
1642 LRL25 = Z(WW) - (.5*1/10^SENS)
1644 LRL75 = Z(VV) - (.5*1/10^SENS)
1650 MD1=(1/10^SENS)*(Q-CUM(YY-1))/G(YY)
1652 P251 = (1/10^SENS)*(Q25 - CUM(WW - 1))/G(WW)
1654 P751 = (1/10^SENS)*(Q75 - CUM(VV - 1))/G(VV)
1660 PRINT
1670 LRL2= Z(YY-1) + (.5*1/10^SENS)
1672 LRL25 = Z(WW - 1) + (.5*1/10^SENS)
```

```
1674 LRL75 = Z(VV - 1) + (.5*1/10^SENS)
1680 MD2 =MD1 + LRL2
1682 P25 = P251 + LRL25
1684 P75 = P751 + LRL75
1690 PRINT "THE MEDIAN EQUALS";  MD2
1692 PRINT "THE 25TH PERCENTILE OF REC. STANDARDS EQUALS";P25
1694 PRINT "THE 75TH PERCENTILE OF REC. STANDARDS EQUALS";P75
1700 LPRINT "THE MEDIAN EQUALS"; MD2
1702 LPRINT
1704 LPRINT "THE 25TH PERCENTILE OF REC. STANDARDS EQUALS";P25
1706 LPRINT
1708 LPRINT "THE 75TH PERCENTILE OF REC. STANDARDS EQUALS";P75
1710 LPRINT
1720 LPRINT
1730 LPRINT
1740 GOTO 1980
1750 INPUT "ENTER DATA FILE NAME AND EXTENSION IN THE FORM
     FILE.EXT";FILE$
1760 INPUT "ENTER THE NUMBER OF TEST ITEMS";NI
1770 OPEN "I",#1,FILE$
1780 FOR I=1 TO N
1790 INPUT#1,SS$(I),ZZ(I)
1800 FOR J=1 TO NI
1810 INPUT#1,X(I,J)
1820 NEXT J
1830 INPUT#1,STAN(I),P(I),PUCO(I)
1840 NEXT I
1850 INPUT "HOW MANY ITEMS ARE BELOW THE RELEVANCE THRESHOLD";NBEL
1860 GOTO 560
1870 FOR I=1 TO NBEL
1880 INPUT "ENTER THE INDEX OF AN ITEM BELOW THE RELEVANCE THRESHOLD";
     IND(I)
1890 NEXT I
1900 FOR I = 1 TO N
1910 FOR J = 1 TO NBEL
1920 K = IND(J)
1930 IF X(I,K)=0 THEN X(I,K)=MEANN(K)
1940 STAN(I) = STAN(I) - X(I,K)
1950 NEXT J
1960 NEXT I
1970 GOTO 380
1980 REM THIS SEGMENT COMPUTES EBEL'S RELIABILITY
1990 REM OF AVERAGE RECOMMENDED STANDARDS
2000 NSUMSQ=0
2010 NSUMST=0
2020 NPROD=0
2030 FOR I = 1 TO N
2040 IF STAN(I)>55 GOTO 2060
2050 NSUMSQ = NSUMSQ + (STAN(I))*(STAN(I))
2060 NEXT I
2070 FOR I = 1 TO N
2080 IF STAN(I)>55 GOTO 2100
2090 NSUMST = NSUMST + STAN(I)
2100 NEXT I
2110 NPROD = (NSUMST)*(NSUMST)/((NI)*(NS))
```

```
2120 SSRAT = (NSUMSQ/NI) - NPROD
2130 PRINT "SUM OF SQUARES FOR JUDGES IS"; SSRAT
2140 LPRINT
2150 LPRINT "SUM OF SQUARES FOR JUDGES IS";SSRAT
2160 LPRINT
2170 SUMIT = 0
2180 FOR J = 1 TO NI
2190 SUMIT = SUMIT + (MEANN(J)*N)*(MEANN(J)*N)
2200 NEXT J
2210 SSITEM = (SUMIT/N) - NPROD
2220 PRINT
2230 PRINT "SUM OF SQUARES FOR ITEMS IS"; SSITEM
2240 LPRINT "SUM OF SQUARES FOR ITEMS IS";SSITEM
2250 LPRINT
2260 SSTOT = 0
2270 FOR I = 1 TO N
2280 FOR J = 1 TO NI
2290 SSTOT = SSTOT + (X(I,J))^2
2300 NEXT J
2310 NEXT I
2320 SSTOT = SSTOT - NPROD
2330 SSEROR = SSTOT - SSITEM - SSRAT
2340 PRINT
2350 PRINT "SUM OF SQUARES TOTAL IS";SSTOT
2360 LPRINT "SUM OF SQUARES TOTAL IS";SSTOT
2370 LPRINT
2380 PRINT "SUM OF SQUARES FOR ERROR IS";SSEROR
2390 LPRINT "SUM OF SQUARES FOR ERROR IS";SSEROR
2400 PRINT
2410 LPRINT
2420 MSITEM = SSITEM/(NI - 1)
2430 MSEROR = SSEROR/((NI - 1)*(NS - 1))
2440 PRINT
2450 PRINT "THE MEAN SQUARE FOR ITEMS IS";MSITEM
2460 LPRINT "THE MEAN SQUARE FOR ITEMS IS";MSITEM
2470 LPRINT
2480 PRINT "THE MEAN SQUARE FOR ERROR IS";MSEROR
2490 LPRINT "THE MEAN SQUARE FOR ERROR IS";MSEROR
2500 LPRINT
2510 PRINT
2520 AVGREL = (MSITEM - MSEROR)/MSITEM
2530 PRINT "THE RELIABILITY OF AVERAGE RECOMMENDED STANDARD IS";
     AVGREL
2540 LPRINT "THE RELIABILITY OF AVERAGE RECOMMENDED STANDARD IS";
     AVGREL
2550 END
```

Once a recommended standard had been computed for each participant who provided judgments on each multiple-choice subtest, the next step in the analysis was to compute various descriptive statistics for the distributions of recommended standards. The STANDESC.BAS computer program was used to compute the following statistics: (1) the minimum recommended standard, (2) the 25th percentile of the distribution of recommneded standards, (3) the 50th percentile of the distribution of recommended standards, (4) the 75th percentile of the distribution of recommended standards, (5) the maximum recommended standard, (6) the mean recommended standard, (7) the variance of recommended standards, (8) the standard deviation of recommended standards, (9) the standard error of the mean recommended standards, and (10) Ebel`s (1951) estimated reliability of the mean recommended standard. Each of these statistics was computed separately for public school teachers and college and university faculty, and for recommendations provided during the first and the second standard-setting sessions.

Analyses Used with the Writing Essay Subtest. Participants who provided recommendations on an appropriate standard for the Writing Essay subtest followed markedly different procedures during the first and the last two judgment sessions. All procedures were replicated, using two distinct sets of essays. The essays used were sampled from a population produced during the April, 1983 administration of the NTE General Knowledge and Communication Skills tests to a sample of students in North Carolina. Each of the essays produced by a North Carolina student was scored by ETS raters, using a scale that ranged from two to twelve. Each set of essays used in the standard-setting sessions contained one essay at each possible score level.

During the first session, each participant was asked to judge each of eleven essays as "acceptable" or "unacceptable." The mechanics of the judgment procedure are described in detail in the script provided in Section II. C., above. A small section of that script is reproduced here, to indicate the nature of the resulting data. Each participant was told:

"Before you make a judgment about the acceptability or unacceptability of any essay, read the entire set in the order given, so that you can become familiar with the entire range of students` responses. After you have read each essay, then go back to the first one, and, considering the circumstances under which it was written, darken the bubble on the "A" column, to the left of [that essay`s] code name, if you want to indicate that the essay:

"Is acceptable under your definition of the minimum standard of writing quality you would require of a beginning teacher education student.

"Darken the bubble in the 'U' column if you want to indicate that the essay:

"Is unacceptable under your definition of a minimum standard of writing quality."

During the first standard-setting session, participants assigned to the Essay subtest provided a binary decision on the acceptability of each essay in each of two sets.

Data from the first standard-setting session were analyzed by computing, for each participant, the minimum of the ETS-assigned scores on the essays that participant judged to be acceptable. The resulting minima were treated as the Essay subtest standards recommended by the participants. Various descriptive statistics were then computed for the distribution of recommended standards, including (1) the minimum, (2) the 25th percentile, (3) the 50th percentile, (4) the 75th percentile, (5) the maximum, (6) the variance, (7) the standard deviation, (8) the mean, and (9) the standard error of the mean. These statistics were used to prepare a variety of tables and graphs.

Prior to the second standard-setting session, participants assigned to the Essay subtest were told the scores that the ETS readers had assigned to each essay. They were then asked to reconsider their initial recommenations by recording the ETS-assigned scores of all essays they (the participants) judged to be acceptable, and the ETS-assigned scores of all essays they judged to be unacceptable. Participants were also asked to record the highest ETS-assigned score among they essays they judged to be acceptable and the lowest ETS-assigned score among the essays they judged to be unacceptable. The same data-recording procedure was used during the third judgment session. Once again, details of the instructions and procedures used are described in the scripts contained in Section II. C., above. Identical procedures were used for both sets of eleven essays.

Data from the second and third judgment sessions were analyzed in several ways. Ultimately, the standard recommended by each participant in response to a given set of essays was taken to be the lowest ETS-assigned score, among the essays that participant found to be acceptable. The descriptive statistics computed for the standards recommended during the first judgment session were also computed for standards recommended during the second and third judgment sessions. A special computer program, titled STANESS.BAS, was written in the BASIC compiler language to analyze these data. It is listed below:

```
10 REM ***************************************************************:
20 REM                        STANESS.BAS
30 REM A PROGRAM TO COMPUTE DESCRIPTIVE STATISTICS ON DATA
40 REM FROM THE NTE WRITING ESSAY TEST -- PHASE III OF THE QUALITY
50 REM ASSURANCE PROGRAM.
60 REM
70 REM WRITTEN BY                  RICHARD M. JAEGER
80 REM LAST REVISED AT 8:20 P.M. ON 13 JANUARY 1984
90 REM
100 REM ************************************************************
110 DIM SS$(60),DATE(60),SCHL(60),AU(5,60,11),SESS(2,60)
120 DIM ALOW(3,60),UHIGH(3,60),INLOW(1,11),INHIGH(1,11),ESS1(60)
130 DIM ESS2(60),ESS3(60),Y(60),F(60),CUM(60),G(60),Z(60)
140 PRINT
150 PRINT "BE SURE THAT THE PRINTER IS CONNECTED AND ON!!!!"
160 REM THIS PROGRAM COMPUTES BASIC DESCRIPTIVE STATISTICS INCLUDING:
170 REM                MEANS
180 REM                MEDIANS
190 REM                25TH PERCENTILES
200 REM                75TH PERCENTILES
210 REM                STANDARD DEVIATIONS
220 REM                STANDARD ERRORS OF THE MEAN
230 REM                VARIANCES
240 REM                ERROR VARIANCES OF THE MEAN
250 REM                MAXIMUM AND MINIMUM SCORES
260 REM
270 PRINT "INPUT A HEADING -- NOT EXCEEDING 60 CHARACTERS"
280 INPUT H$
290 PRINT
300 PRINT H$
310 LPRINT H$
320 LPRINT
330 LPRINT
340 PRINT
350 PRINT "INPUT THE SAMPLE SIZE"
360 INPUT N
370 PRINT
380 PRINT "THE SAMPLE SIZE IS";N
390 INPUT "ENTER DATA FILE NAME AND EXTENSION IN THE FORM
    FILE.EXT";FILE$
400 OPEN "I",#1,FILE$
410 FOR I=1 TO N
420 INPUT#1,SS$(I),DATE(I),SCHL(I)
430 FOR J=1 TO 11
440 INPUT#1,AU(1,I,J)
450 NEXT J
460 INPUT#1,SESS(1,I)
470 FOR J=1 TO 11
```

```
480 INPUT#1,AU(2,I,J)
490 NEXT J
500 INPUT#1,ALOW(2,I)
510 FOR J=1 TO 11
520 INPUT#1,AU(3,I,J)
530 NEXT J
540 INPUT#1,UHIGH(2,I),SESS(2,I)
550 FOR J=1 TO 11
560 INPUT#1,AU(4,I,J)
570 NEXT J
580 INPUT#1,ALOW(3,I)
590 FOR J=1 TO 11
600 INPUT#1,AU(5,I,J)
610 NEXT J
620 INPUT#1,UHIGH(3,I),JUNK1,JUNK2,JUNK3,JUNK4
630 NEXT I
640 FOR I=1 TO N
650 FOR J=1 TO 11
660 INLOW(1,J)=0
670 INHIGH(1,J)=0
680 IF AU(1,I,J)=5 THEN INLOW(1,J)=J+1
690 IF AU(1,I,J)=6 THEN INHIGH(1,J)=J+1
700 NEXT J
710 ALOW(1,I)=12
720 UHIGH(1,I)=0
730 FOR J=1 TO 11
740 IF INLOW(1,J)=0 GOTO 760
750 IF INLOW(1,J)<ALOW(1,I) THEN ALOW(1,I)=INLOW(1,J)
760 NEXT J
770 FOR J=1 TO 11
780 IF INHIGH(1,J)=0 GOTO 800
790 IF INHIGH(1,J)>UHIGH(1,I) THEN UHIGH(1,I)=INHIGH(1,J)
800 NEXT J
810 PRINT "THE LOWEST ACCEPTABLE ESSAY FROM SESSION 1 IS";ALOW(1,I)
820 PRINT
830 PRINT "THE HIGHEST UNACCEPTABLE ESSAY FROM SESSION 1 IS";
    UHIGH(1,I)
840 ESS1(I)=(ALOW(1,I)+UHIGH(1,I)+1)/2
850 NEXT I
860 FOR I=1 TO N
870 ESS2(I)=(ALOW(2,I)+UHIGH(2,I)+1)/2
880 ESS3(I)=(ALOW(3,I)+UHIGH(3,I)+1)/2
890 NEXT I
900 SUM1=0
910 SUM2=0
920 SUM3=0
930 SUMSQ1=0
940 SUMSQ2=0
950 SUMSQ3=0
960 FOR I=1 TO N
970 SUM1=SUM1+ESS1(I)
980 SUM2=SUM2+ESS2(I)
990 SUM3=SUM3+ESS3(I)
1000 SUMSQ1=SUMSQ1+(ESS1(I))^2
1010 SUMSQ2=SUMSQ2+(ESS2(I))^2
```

```
1020 SUMSQ3=SUMSQ3+(ESS3(I))^2
1030 NEXT I
1040 MEAN1=SUM1/N
1050 MEAN2=SUM2/N
1060 MEAN3=SUM3/N
1070 VAR1=(SUMSQ1-(SUM1^2/N))/(N-1)
1080 VAR2=(SUMSQ2-(SUM2^2/N))/(N-1)
1090 VAR3=(SUMSQ3-(SUM3^2/N))/(N-1)
1100 STD1=SQR(VAR1)
1110 STD2=SQR(VAR2)
1120 STD3=SQR(VAR3)
1130 PRINT "THE MEAN RECOMMENDED STANDARD FOR SESSION 1 IS";MEAN1
1140 PRINT "THE MEAN RECOMMENDED STANDARD FOR SESSION 2 IS";MEAN2
1150 PRINT "THE MEAN RECOMMENDED STANDARD FOR SESSION 3 IS";MEAN3
1160 PRINT
1170 LPRINT "THE MEAN RECOMMENDED STANDARD FOR SESSION 1 IS";MEAN1
1180 LPRINT "THE MEAN RECOMMENDED STANDARD FOR SESSION 2 IS";MEAN2
1190 LPRINT "THE MEAN RECOMMENDED STANDARD FOR SESSION 3 IS";MEAN3
1200 LPRINT
1210 PRINT
1220 PRINT "THE STANDARD DEVIATION OF RECOMMENDATIONS FOR
     SESSION 1 IS";STD1
1230 PRINT "THE STANDARD DEVIATION OF RECOMMENDATIONS FOR
     SESSION 2 IS";STD2
1240 PRINT "THE STANDARD DEVIATION OF RECOMMENDATIONS FOR
     SESSION 3 IS";STD3
1250 PRINT
1260 LPRINT "THE STANDARD DEVIATION OF RECOMMENDATIONS FOR
     SESSION 1 IS";STD1
1270 LPRINT "THE STANDARD DEVIATION OF RECOMMENDATIONS FOR
     SESSION 2 IS";STD2
1280 LPRINT "THE STANDARD DEVIATION OF RECOMMENDATIONS FOR
     SESSION 3 IS";STD3
1290 LPRINT
1300 REM THIS BLOCK COMPUTES STANDARD ERRORS OF THE MEAN
1310 SEM1=STD1/SQR(N)
1320 SEM2=STD2/SQR(N)
1330 SEM3=STD3/SQR(N)
1340 REM THIS BLOCK COMPUTES VARIANCE ERRORS OF THE MEAN
1350 ERV1=SEM1^2
1360 ERV2=SEM2^2
1370 ERV3=SEM3^2
1380 PRINT "THE STANDARD ERROR OF THE MEAN FOR SESSION 1 IS";SEM1
1390 PRINT "THE STANDARD ERROR OF THE MEAN FOR SESSION 2 IS";SEM2
1400 PRINT "THE STANDARD ERROR OF THE MEAN FOR SESSION 3 IS";SEM3
1410 PRINT
1420 LPRINT "THE STANDARD ERROR OF THE MEAN FOR SESSION 1 IS";SEM1
1430 LPRINT "THE STANDARD ERROR OF THE MEAN FOR SESSION 2 IS";SEM2
1440 LPRINT "THE STANDARD ERROR OF THE MEAN FOR SESSION 3 IS";SEM3
1450 LPRINT
1460 PRINT "THE VARIANCE ERROR OF THE MEAN FOR SESSION 1 IS";ERV1
1470 PRINT "THE VARIANCE ERROR OF THE MEAN FOR SESSION 2 IS";ERV2
1480 PRINT "THE VARIANCE ERROR OF THE MEAN FOR SESSION 3 IS";ERV3
1490 PRINT
1500 LPRINT "THE VARIANCE ERROR OF THE MEAN FOR SESSION 1 IS";ERV1
```

```
1510 LPRINT "THE VARIANCE ERROR OF THE MEAN FOR SESSION 2 IS";ERV2
1520 LPRINT "THE VARIANCE ERROR OF THE MEAN FOR SESSION 3 IS";ERV3
1530 LPRINT
1540 FOR I=1 TO N
1550 Y(I)=ESS1(I)
1560 NEXT I
1570 SESCT=1
1580 FLIPS=1
1590 WHILE FLIPS
1600 FLIPS=0
1610 FOR I=1 TO N-1
1620 IF Y(I)>Y(I+1) THEN SWAP Y(I),Y(I+1):FLIPS=1
1630 NEXT I
1640 WEND
1650 MIN=Y(1)
1660 MAX=Y(N)
1670 RANGE=MAX-MIN
1680 PRINT "THE SMALLEST AND LARGEST RECOMMENDED STANDARDS FOR
     SESSION";SESCT;" ARE"
1690 PRINT "    ";MIN;"  AND  ";MAX
1700 PRINT
1710 LPRINT "THE SMALLEST AND LARGEST RECOMMENDED STANDARDS FOR
     SESSION";SESCT;"ARE"
1720 LPRINT "   ";MIN;"  AND  ";MAX
1730 LPRINT
1740 INCL = RANGE+1
1750 PRINT "THE EXCLUSIVE RANGE FOR SESSION";SESCT;"IS";RANGE
1760 PRINT "THE INCLUSIVE RANGE FOR SESSION";SESCT;"IS";INCL
1770 PRINT
1780 LPRINT "THE EXCLUSIVE RANGE FOR SESSION";SESCT;"IS";RANGE
1790 LPRINT "THE INCLUSIVE RANGE FOR SESSION";SESCT;"IS";INCL
1800 FOR I=1 TO N
1810 F(I)=1
1820 NEXT I
1830 FOR I=1 TO N
1840 FOR J=I+1 TO N
1850 IF Y(J)=Y(I) AND F(J)>0 THEN F(I)=F(I)+1
1860 IF Y(J)=Y(I) AND F(J)>0 THEN F(J)=F(J)-1
1870 NEXT J
1880 NEXT I
1890 FOR I=1 TO N
1900 IF F(I)=0 THEN Y(I)=0
1910 NEXT I
1920 K=1
1930 FOR I=1 TO N
1940 IF Y(I)>0 THEN Z(K)=Y(I)
1950 IF Y(I)>0 THEN G(K)=F(I)
1960 IF Y(I)>0 THEN K=K+1
1970 NEXT I
1980 FOR L=1 TO K-1
1990 NEXT L
2000 CUM(1)=G(1)
2010 FOR L=2 TO K-1
2020 CUM(L)=CUM(L-1)+G(L)
2030 NEXT L
```

```
2040 Q=N/2
2050 QQ=N/4
2060 QQQ=3*(N/4)
2070 YY=1
2080 YYY=1
2090 YYYY=1
2100 FOR L=1 TO K-1
2110 IF CUM(L)<=Q THEN YY=YY+1
2120 IF CUM(L)<=QQ THEN YYY=YYY+1
2130 IF CUM(L)<=QQQ THEN YYYY=YYYY+1
2140 NEXT L
2150 LRL50=Z(YY)-.5
2160 LRL25=Z(YYY)-.5
2170 LRL75=Z(YYYY)-.5
2180 MD1=(Q-CUM(YY-1))/G(YY)
2190 LRL502=Z(YY-1)+.5
2200 MD2=MD1+LRL502
2210 PRINT "THE MEDIAN FOR SESSION";SESCT;"IS";MD2
2220 P251=(QQ-CUM(YYY-1))/G(YYY)
2230 LRL252=Z(YYY-1)+.5
2240 P252=P251+LRL252
2250 P751=(QQQ-CUM(YYYY-1))/G(YYYY)
2260 LRL752=Z(YYYY-1)+.5
2270 P752=P751+LRL752
2280 PRINT "THE 25TH PERCENTILE FOR SESSION";SESCT;"IS";P252
2290 PRINT "THE 75TH PERCENTILE FOR SESSION";SESCT;"IS";P752
2300 PRINT
2310 LPRINT "THE MEDIAN FOR SESSION";SESCT;"IS";MD2
2320 LPRINT "THE 25TH PERCENTILE FOR SESSION";SESCT;"IS";P252
2330 LPRINT "THE 75TH PERCENTILE FOR SESSION";SESCT;"IS";P752
2340 IF SESCT=2 GOTO 2410
2350 IF SESCT=3 GOTO 2460
2360 SESCT=SESCT+1
2370 FOR I=1 TO N
2380 Y(I)=ESS2(I)
2390 NEXT I
2400 GOTO 1580
2410 SESCT=SESCT+1
2420 FOR I=1 TO N
2430 Y(I)=ESS3(I)
2440 NEXT I
2450 GOTO 1580
2460 END
```

The statistics computed for each set of essays were averaged to develop more reliable recommendations on an appropriate standard for the Essay subtest.

Estimation of Subtest Reliability. The General Knowledge and Communications Skills Tests of the NTE were administered to a sample of approximately 800 students in North Carolina in April of 1983. Data from this test administration were used to estimate the reliability of each subtest and the standard error of measurement on each subtest. The Reliability procedure of the Statistical Package for the Social Sciences, Version-X (SPSS^{x}) was used to compute these statistics for each of the multiple-choice subtests.

Cronbach's alpha was used as an estimate of the reliability of each multiple-choice subtest. The usual classical test theory formulation was used to estimate the standard error of measurement on each of these subtests, based on Cronbach's alpha and the standard deviation of examinees' raw scores on the subtest.

The data tape provided by Educational Testing Service listed the score assigned to each examinee's essay, on a scale of one to six, by each of two raters. These scores were used to compute an inter-scorer reliability coefficient, which was then used as an estimate of the half-scale reliability of the Essay test. The Spearman-Brown Prophecy Formula was used to adjust this reliability estimate to the two to twelve scale used in the final ETS scoring of the essays.

The standard error of measurement of the Essay test was estimated by using the usual classical test theory formulation, based on the estimated reliability and the standard deviation of esaminees' essay scores on the ETS scale of two to twelve.

Estimation of Total Standard Errors. A "total" standard error was estimated for each subtest, by combining the standard error of measurement of that subtest with the standard error of the mean recommended standard on that subtest. This total standard error was thus sensitive to two important error components that might influence recommended standards: random errors due to the selection of specific items for the subtest, and random errors due to the selection of a particular sample of participants for the standard-setting process.

The total standard error for a subtest was calculated as the square root of the sum of the <u>squares</u> of that subtest`s standard error of measurement and the standard error of the mean recommended standard for that subtest.

<u>Computation of Recommended Standards for the Tests</u>. Once recommended standards had been developed for each subtest, it was necessary to combine these recommendations so as to produce recommended standards for the NTE General Knowledge and Communication Skills Tests. When Educational Testing Service computes examinees` scores on these tests, they are based on weighted averages of performances on each of the four subtests that compose each test. The formulas used to compute recommended standards on the tests were identical to those used by ETS in computing examinees` "raw" (number right) and scaled scores.

For the General Knowledge Test, recommended raw-score standards were computed from recommended raw-score standards on the Science (SCI), Mathematics (MAT), Literature and Fine Arts (LIT), and Social Studies (SOC) subtests by using the following formula:

$$GKR = 1.1688\ SOC + 1.2243\ MAT + 1.0\ LIT + 1.1666\ SCI\ .$$

Raw-score recommended standards were then converted to NTE General Knowledge scaled scores through the linear transformation:

$$GKSCA = 0.6618\ GKR + 600.0\ .$$

For the Communication Skills Test, raw-score recommended standards were computed from recommended raw-score standards on the Listening (LIS), Reading (RED), Writing multiple-choice (WRI) and Writing Essay (ESS) subtests by using the following formula:

$$CSR = 2.3624\ LIS + 2.6313\ RED + 1.0\ WRI + 3.8581\ ESS\ .$$

Raw-score recommended standards were then converted to NTE Communication Skills scaled scores through the linear transformation:

$$CSSCA = 0.3396\ CSR + 600.0\ .$$

Adjustment for Errors of Measurement and Estimation. Were recommended standards to be adopted directly, fifty percent of examinees with true scores equal to the recommended standards would fail each test, just because of errors of measurement and/or errors of estimation. To protect against this possibility, it is necessary to adjust the recommended standards by using estimated total errors.

Error adjustments were computed using two different measurement models. One type of adjustment assumes that examinees are to be screened on the basis of their performances on each subtest. In this model, the recommended standard on each subtest is reduced by a multiple of the total standard error on that subtest. The other type of adjustment assumes that examinees are to be screened on the basis of their performances on entire tests (i.e., the NTE General Knowledge Test and the NTE Communication Skills Test), and adjusts recommended standards by subtracting a multiple of the total standard error on each test. The latter procedure results in somewhat more stringent standards, but is consistent with the selection model currently envisioned by the State Board of Education of North Carolina. Both types of error models were applied to the standard-setting data resulting from this study.

Recommended standards on the General Knowledge and Communication Skills Tests, and adjustments for errors of measurement and estimation using the subtest-by-subtest adjustment model were determined by using a specially developed computer program, written in the BASIC compiler language. This program, titled SCALSTAN.BAS, is listed below:

```
10 REM THIS IS A PROGRAM FOR COMPUTING SCALED-SCORE STANDARDS ON THE
20 REM NATIONAL TEACHER EXAMINATIONS -- GENERAL KNOWLEDGE TEST AND
30 REM COMMUNICATION SKILLS TEST.
40 REM
50 REM IT ASSUMES INPUT OF RAW-SCORE STANDARDS AND VARIOUS STANDARD
60 REM ERROR ESTIMATES ON ALL SUBTESTS.
70 REM
80 REM THE PROGRAM WAS WRITTEN BY RICHARD M. JAEGER ON 11 JANUARY
90 REM
100 REM                AT 4:50 P.M.
110 REM
120 PRINT "BE SURE THAT THE PRINTER IS CONNECTED AND ON"
130 PRINT "OTHERWISE THIS PROGRAM WILL NOT RUN!!!!"
140 PRINT
150 INPUT "TYPE A HEADING, NOT TO EXCEED 60 CHARACTERS";HEAD$
160 PRINT
170 PRINT HEAD$
180 LPRINT HEAD$
190 PRINT
200 LPRINT
210 INPUT "TYPE THE RAW-SCORE STANDARD ON THE SOCIAL STUDIES TEST";
    SOC
220 INPUT "TYPE THE RAW-SCORE STANDARD ON THE MATHEMATICS TEST";
    MAT
230 INPUT "TYPE THE RAW-SCORE STANDARD ON THE LITERATURE-FINE ARTS
    TEST";LIT
240 INPUT "TYPE THE RAW-SCORE STANDARD ON THE SCIENCE TEST";SCI
250 INPUT "TYPE THE RAW-SCORE STANDARD ON THE LISTENING TEST";LIS
260 INPUT "TYPE THE RAW-SCORE STANDARD ON THE READING TEST";REA
270 INPUT "TYPE THE RAW-SCORE STANDARD ON THE WRITING M.C. TEST";WRI
280 INPUT "TYPE THE RAW-SCORE STANDARD ON THE WRITING ESSAY TEST";ESS
```

```
290 PRINT
300 INPUT "TYPE THE STD. ERROR OF MEASUREMENT ON THE SOCIAL STUDIES
    TEST";SEMSOC
310 INPUT "TYPE THE STD. ERROR OF MEASUREMENT ON THE MATHEMATICS TEST";
    SEMMAT
320 INPUT "TYPE THE STD. ERROR OF MEASUREMENT ON THE LITERATURE-FINE
    ARTS TEST";SEMLIT
330 INPUT "TYPE THE STD. ERROR OF MEASUREMENT ON THE SCIENCE TEST";
    SEMSCI
340 INPUT "TYPE THE STD. ERROR OF MEASUREMENT ON THE LISTENING TEST";
    SEMLIS
350 INPUT "TYPE THE STD. ERROR OF MEASUREMENT ON THE READING TEST";
    SEMREA
360 INPUT "TYPE THE STD. ERROR OF MEASUREMENT ON THE WRITING M.C.
    TEST";SEMWRI
370 INPUT "TYPE THE STD. ERROR OF MEASUREMENT ON THE WRITING ESSAY
    TEST";SEMESS
380 PRINT
390 INPUT "TYPE THE STANDARD ERROR OF THE MEAN ON THE SOCIAL STUDIES
    TEST";SDSOC
400 INPUT "TYPE THE STANDARD ERROR OF THE MEAN ON THE MATHEMATICS
    TEST";SDMAT
410 INPUT "TYPE THE STANDARD ERROR OF THE MEAN ON THE LITERATURE F.A.
    TEST";SDLIT
420 INPUT "TYPE THE STANDARD ERROR OF THE MEAN ON THE SCIENCE
    TEST";SDSCI
430 INPUT "TYPE THE STANDARD ERROR OF THE MEAN ON THE LISTENING
    TEST";SDLIS
440 INPUT "TYPE THE STANDARD ERROR OF THE MEAN ON THE READING TEST";
    SDREA
450 INPUT "TYPE THE STANDARD ERROR OF THE MEAN ON THE WRITING M.C.
    TEST";SDWRI
460 INPUT "TYPE THE STANDARD ERROR OF THE MEAN ON THE WRITING ESSAY
    TEST";SDESS
470 PRINT
480 GKR=1.1688*SOC+1.2243*MAT+1!*LIT+1.1666*SCI
490 CSR=2.3624*LIS+2.6313*REA+1!*WRI+3.8581*ESS
500 GKSCA=.6618*GKR+600!
510 CSSCA=.3396*CSR+600!
520 PRINT
530 PRINT "THE GENERAL KNOWLEDGE MEAN RAW STANDARD IS";GKR
540 LPRINT "THE GENERAL KNOWLEDGE MEAN RAW STANDARD IS";GKR
550 LPRINT
560 PRINT
570 PRINT "THE COMMUNICATION SKILLS MEAN RAW STANDARS IS";CSR
580 LPRINT "THE COMMUNICATION SKILLS MEAN RAW STANDARD IS";CSR
590 SOCERR=SQR(SEMSOC^2 + SDSOC^2)
600 MATERR=SQR(SEMMAT^2 + SDMAT^2)
610 LITERR=SQR(SEMLIT^2 + SDLIT^2)
620 SCIERR=SQR(SEMSCI^2 + SDSCI^2)
630 LISERR=SQR(SEMLIS^2 + SDLIS^2)
640 REDERR=SQR(SEMREA^2 + SDREA^2)
650 WRIERR=SQR(SEMWRI^2 + SDWRI^2)
660 ESSERR=SQR(SEMESS^2 + SDESS^2)
670 PRINT
```

```
680 LPRINT
690 PRINT "THE SCALED SCORE MEAN RECOMM. STANDARD ON GEN. KNOW.
    IS";GKSCA
700 LPRINT "THE SCALED SCORE MEAN. RECOMM. STANDARD ON GEN. KNOW.
    IS";GKSCA
710 LPRINT
720 LPRINT "THE SCALED SCORE MEAN RECOMM. STANDARD ON COMM. SKILLS
    IS";CSSCA
730 PRINT "THE SCALED SCORE MEAN RECOMM. STANDARD ON COMM. SKILLS
    IS";CSSCA
740 PRINT
750 LPRINT
760 LPRINT "THE TOTAL STANDARD ERROR FOR SOCIAL STUDIES IS";SOCERR
770 LPRINT "THE TOTAL STANDARD ERROR FOR MATHEMATICS IS";MATERR
780 LPRINT "THE TOTAL STANDARD ERROR FOR LITERATURE-F.A. IS";LITERR
790 LPRINT "THE TOTAL STANDARD ERROR FOR SCIENCE IS";SCIERR
800 LPRINT "THE TOTAL STANDARD ERROR FOR LISTENING IS";LISERR
810 LPRINT "THE TOTAL STANDARD ERROR FOR READING IS";REDERR
820 LPRINT "THE TOTAL STANDARD ERROR FOR WRITING M.C. IS";WRIERR
830 LPRINT "THE TOTAL STANDARD ERROR FOR WRITING ESSAY IS";ESSERR
840 REM THE NEXT BLOCK CALCULATES RAW SCORE MEAN STANDARDS
850 REM MINUS ONE TOTAL STANDARD ERROR
860 SOCM1=SOC-SOCERR
870 MATM1=SOC-MATERR
880 LITM1=LIT-LITERR
890 SCIM1=SCI-SCIERR
900 LISM1=LIS-LISERR
910 REDM1=REA-REDERR
920 WRIM1=WRI-WRIERR
930 ESSM1=ESS-ESSERR
940 REM THE NEXT BLOCK CALCULATES RAW SCORE MEAN STANDARDS MINUS
950 REM TWO TOTAL STANDARD ERRORS
960 SOCM2=SOC-2*SOCERR
970 MATM2=MAT-2*MATERR
980 LITM2=LIT-2*LITERR
990 SCIM2=SCI-2*SCIERR
1000 LISM2=LIS-2*LISERR
1010 REDM2=REA-2*REDERR
1020 WRIM2=WRI-2*WRIERR
1030 ESSM2=ESS-2*ESSERR
1040 REM THE NEXT BLOCK CALCULATES TOTAL RAW SCORE AND SCALED SCORE
1050 REM STANDARDS BASED ON RECOMMENDED MEANS MINUS ONE STANDARD
     ERROR
1060 GKRM1=1.1688*SOCM1+1.2243*MATM1+1!*LITM1+1.1666*SCIM1
1070 CSRM1=2.3624*LISM1+2.6313*REDM1+1!*WRIM1+3.8581*ESSM1
1080 GKSCAM1=.6618*GKRM1+600!
1090 CSSCAM1=.3396*CSRM1+600!
1100 PRINT
1110 PRINT "ON GEN. KNOW. THE MEAN SCALED SCORE STANDARD MINUS ONE
     STD.ERR. IS";GKSCAM1
1120 LPRINT
1130 LPRINT "ON GEN. KNOW. THE MEAN SCALED SCORE STANDARD MINUS ONE
     STD.ERR. IS";GKSCAM1
1140 PRINT "ON COMM. SKILLS THE MEAN SCALED SCORE STANDARD MINUS ONE
     STD.ERR. IS";CSSCAM1
```

```
1150 LPRINT "ON COMM. SKILLS THE MEAN SCALED SCORE STANDARD MINUS ONE
     STD.ERR. IS";CSSCAM1
1160 GKRM2=1.1688*SOCM2+1.2243*MATM2+1!*LITM2+1.1666*SCIM2
1170 CSRM2=2.3624*LISM2+2.6313*REDM2+1!*WRIM2+3.8581*ESSM2
1180 GKSCAM2=.6618*GKRM2+600!
1190 CSSCAM2=.3396*CSRM2+600!
1200 LPRINT
1210 PRINT "ON GEN. KNOW. THE MEAN SCALED SCORE STANDARD MINUS
     TWO STD.ERR. IS";GKSCAM2
1220 LPRINT "ON GEN. KNOW. THE MEAN SCALED SCORE STANDARD MINUS
     TWO STD.ERR. IS";GKSCAM2
1230 PRINT "ON COMM. SKILLS THE MEAN SCALED SCORE STANDARD MINUS
     TWO STD.ERR. IS";CSSCAM2
1240 LPRINT "ON COMM. SKILLS THE MEAN SCALED SCORE STANDARD MINUS
     TWO STD.ERR. IS";CSSCAM2
1250 REM THE NEXT BLOCK CALCULATES RAW SCORE MEAN STANDARDS MINUS
1260 REM THREE TOTAL STANDARD ERRORS
1270 SOCM3=SOC-3*SOCERR
1280 MATM3=MAT-3*MATERR
1290 LITM3=LIT-3*LITERR
1300 SCIM3=SCI-3*SCIERR
1310 LISM3=LIS-3*LISERR
1320 REDM3=REA-3*REDERR
1330 WRIM3=WRI-3*WRIERR
1340 ESSM3=ESS-3*ESSERR
1350 GKRM3=1.1688*SOCM3+1.2243*MATM3+1!*LITM3+1.1666*SCIM3
1360 CSRM3=2.3624*LISM3+2.6313*REDM3+1!*WRIM3+3.8581*ESSM3
1370 GKSCAM3=.6618*GKRM3+600!
1380 CSSCAM3=.3396*CSRM3+600!
1390 LPRINT
1400 PRINT "ON GEN. KNOW. THE MEAN SCALED SCORE STANDARD MINUS
     THREE STD.ERR. IS";GKSCAM3
1410 LPRINT "ON GEN. KNOW. THE MEAN SCALED SCORE STANDARD MINUS
     THREE STD.ERR. IS";GKSCAM3
1420 PRINT "ON COMM. SKILLS THE MEAN SCALED SCORE STANDARD MINUS
     THREE WO STD.ERR. IS";CSSCAM3
1430 LPRINT "ON COMM. SKILLS THE MEAN SCALED SCORE STANDARD MINUS
     THREE STD.ERR. IS";CSSCAM3
1440 END
```

Another BASIC computer program, titled SQRSTD.BAS, was written to apply the second error adjustment model. In this program, the weighted averages and linear transformations defined above were applied to the total standard error estimates computed for each subtest. Multiples of the resulting total standard errors were then subtracted from the scaled score recommended standards on the General Knowlege and Communication Skills Tests. A listing of this computer program follows:

```
10 REM THIS IS A NEW PROGRAM FOR COMPUTING NTE SCALED SCORES
20 REM ADJUSTED FOR ERRORS OF MEASUREMENT AND ERRORS DUE TO
   SELECTION
30 REM OF JUDGES. IT INCORPORATES A ROOT MEAN SQUARE ERROR MODEL,
40 REM ADJUSTING THE GENERAL KNOWLEDGE RAW SCORE AND THE
   COMMUNICATION
50 REM SKILLS RAW SCORE, INSTEAD OF THE PART SCORES.
60 REM
70 REM IT WAS WRITTEN BY RICHARD M. JAEGER
80 REM ON             7 FEBRUARY 1984
90 REM      AT        4:20 A.M.
100 REM
110 A1 = 1.1688
120 A2 = 1.2243
130 A3 = 1!
140 A4 = 1.1666
150 B1 = 2.3624
160 B2 = 2.6313
170 B3 = 1!
180 B4 = 3.8581
190 PRINT "BE SURE THAT THE PRINTER IS CONNECTED AND ON!!!!"
200 PRINT
210 INPUT "TYPE A HEADING NOT TO EXCEED 60 CHARACTERS";HEAD$
220 PRINT HEAD$
230 LPRINT HEAD$
240 LPRINT
250 INPUT "WHAT IS THE RECOMMENDED STANDARD FOR SOCIAL STUDIES";SOC
260 INPUT "WHAT IS THE RECOMMENDED STANDARD FOR MATHEMATICS";MAT
270 INPUT "WHAT IS THE RECOMMENDED STANDARD FOR LITERATURE/
    FINE ARTS";LIT
280 INPUT "WHAT IS THE RECOMMENDED STANDARD FOR SCIENCE";SCI
290 PRINT
300 PRINT "THE RECOMMENDED STANDARD FOR SOCIAL STUDIES IS";SOC
310 PRINT "THE RECOMMENDED STANDARD FOR MATHEMATICS IS";MAT
320 PRINT "THE RECOMMENDED STANDARD FOR LITERATURE/FINE
    ARTS IS";LIT
330 PRINT "THE RECOMMENDED STANDARD FOR SCIENCE IS";SCI
340 PRINT
350 INPUT "ARE THESE VALUES CORRECT";RESP$
360 IF LEFT$(RESP$,1)<>"Y" GOTO 250
370 PRINT
375 IF PASS=2 GOTO 420
380 INPUT "WHAT IS THE TOTAL ERROR FOR SOCIAL STUDIES";ESOC
390 INPUT "WHAT IS THE TOTAL ERROR FOR MATHEMATICS";EMAT
400 INPUT "WHAT IS THE TOTAL ERROR FOR LITERATURE/FINE ARTS";ELIT
410 INPUT "WHAT IS THE TOTAL ERROR FOR SCIENCE";ESCI
420 PRINT
430 PRINT "THE TOTAL ERROR FOR SOCIAL STUDIES IS";ESOC
440 PRINT "THE TOTAL ERROR FOR MATHEMATICS IS";EMAT
450 PRINT "THE TOTAL ERROR FOR LITERATURE/FINE ARTS IS";ELIT
460 PRINT "THE TOTAL ERROR FOR SCIENCE IS";ESCI
470 PRINT
480 INPUT "ARE THESE VALUES CORRECT";RESP2$
490 IF LEFT$(RESP2$,1)<>"Y" GOTO 380
```

```
500 PRINT
510 INPUT "WHAT IS THE RECOMMENDED STANDARD FOR LISTENING";LIS
520 INPUT "WHAT IS THE RECOMMENDED STANDARD FOR READING";RED
530 INPUT "WHAT IS THE RECOMMENDED STANDARD FOR WRITING MC";WRI
540 INPUT "WHAT IS THE RECOMMENDED STANDARD FOR WRITING ESSAY";ESS
550 PRINT
560 PRINT "THE RECOMMENDED STANDARD FOR LISTENING IS";LIS
570 PRINT "THE RECOMMENDED STANDARD FOR READING IS";RED
580 PRINT "THE RECOMMENDED STANDARD FOR WRITING MC IS";WRI
590 PRINT "tHE RECOMMENDED STANDARD FOR WRITING ESSAY IS";ESS
600 PRINT
610 INPUT "ARE THESE VALUES CORRECT";RESP3$
620 IF LEFT$(RESP3$,1)<>"Y" GOTO 510
625 IF PASS=2 GOTO 670
630 INPUT "WHAT IS THE TOTAL ERROR FOR LISTENING";ELIS
640 INPUT "WHAT IS THE TOTAL ERROR FOR READING";ERED
650 INPUT "WHAT IS THE TOTAL ERROR FOR WRITING MC";EWRI
660 INPUT "WHAT IS THE TOTAL ERROR FOR WRITING ESSAY";EESS
670 PRINT
680 PRINT "THE TOTAL ERROR FOR LISTENING IS";ELIS
690 PRINT "THE TOTAL ERROR FOR READING IS";ERED
700 PRINT "THE TOTAL ERROR FOR WRITING MC IS";EWRI
710 PRINT "THE TOTAL ERROR FOR WRITING ESSAY IS";EESS
720 PRINT
730 INPUT "ARE THESE VALUES CORRECT";RESP4$
740 IF LEFT$(RESP4$,1)<>"Y" GOTO 630
750 GKR = A1*SOC + A2*MAT + A3*LIT + A4*SCI
760 PRINT
770 CSR = B1*LIS + B2*RED + B3*WRI + B4*ESS
780 PRINT "THE RAW RECOMMENDED STANDARD ON GENERAL KNOWLEDGE
    IS";GKR
790 PRINT "THE RAW RECOMMENDED STANDARD ON COMMUNICATION SKILLS
    IS";CSR
800 PRINT
810 LPRINT "THE RAW RECOMMENDED STANDARD ON GENERAL KNOWLEDGE
    IS";GKR
820 LPRINT "THE RAW RECOMMENDED STANDARD ON COMMUNICATION SKILLS
    IS";CSR
830 EGKR = SQR((A1^2)*(ESOC^2)+(A2^2)*(EMAT^2)+(A3^2)*(ELIT^2)+
    (A4^2)*(ESCI^2))
840 ECSR = SQR((B1^2)*(ELIS^2)+(B2^2)*(ERED^2)+(B3^2)*(EWRI^2)+
    (B4^2)*(EESS^2))
850 PRINT "THE STANDARD ERROR OF GENERAL KNOWLEDGE - RAW SCORE
    IS";EGKR
860 PRINT "THE STANDARD ERROR OF COMMUNICATION SKILLS - RAW SCORE
    IS";ECSR
870 PRINT
880 LPRINT
890 LPRINT "THE STANDARD ERROR OF GENERAL KNOWLEDGE RAW SCORES IS";
    EGKR
900 LPRINT "THE STANDARD ERROR OF COMMUNICATION SKILLS RAW SCORES
    IS";ECSR
910 LPRINT
920 GKSCA = .6618*GKR + 600
930 CSSCA = .3396*CSR + 600
```

```
940 PRINT "THE RECOMMENDED SCALED SCORE STANDARD ON GENERAL KNOWLEDGE
    IS";GKSCA
950 PRINT "THE RECOMMENDED SCALED SCORE STANDARD ON COMMUNICATION
    SKILLS IS";CSSCA
960 PRINT
990 LPRINT
1000 GKSCA1 = .6618*(GKR - EGKR) + 600
1010 CSSCA1 = .3396*(CSR - ECSR) + 600
1020 GKSCA2 = .6618*(GKR - 2*EGKR) + 600
1030 CSSCA2 = .3396*(CSR - 2*ECSR) + 600
1040 GKSCA3 = .6618*(GKR - 3*EGKR) + 600
1050 CSSCA3 = .3396*(CSR - 3*ECSR) + 600
1060 PRINT "FOLLOWING ARE SCALED SCORE RECOMMENDATIONS FOR GENERAL
     KNOWLEDGE"
1070 PRINT "RECOMMENDATION     MINUS 1 S.E.     MINUS 2 S.E.
     MINUS 3 S.E."
1080 PRINT GKSCA;"     ";GKSCA1;"     ";GKSCA2;"     ";GKSCA3
1090 LPRINT "FOLLOWING ARE SCALED SCORE RECOMMENDATIONS FOR GENERAL
     KNOWLEDGE"
1100 LPRINT "RECOMMENDATION     MINUS 1 S.E.     MINUS 2 S.E.
     MINUS 3 S.E."
1110 LPRINT "     ";GKSCA;"     ";GKSCA1;"          ";GKSCA2;"         "
     ;GKSCA3
1120 LPRINT
1130 PRINT
1140 PRINT "FOLLOWING ARE RECOMMENDATIONS FOR COMMUNICATION SKILLS"
1150 PRINT "RECOMMENDATION     MINUS 1 S.E.     MINUS 2 S.E.
     MINUS 3 S.E."
1160 PRINT CSSCA;"     ";CSSCA1;"     ";CSSCA2;"     ";CSSCA3
1170 LPRINT "FOLLOWING ARE RECOMMENDATIONS FOR COMMUNICATION SKILLS"
1180 LPRINT "RECOMMENDATION     MINUS 1 S.E.     MINUS 2 S.E.
     MINUS 3 S.E."
1190 LPRINT "     ";CSSCA;"     ";CSSCA1;"          ";CSSCA2;"
     ";CSSCA3
1195 PRINT
1200 INPUT "DO YOU WANT TO RUN THE PROGRAM WITH DIFFERENT
     STANDARDS";ANS$

1205 PASS = 2
1210 IF LEFT$(ANS$,1)="Y" GOTO 210
1220 END
```

Final Recommended Standards. The standards recommended to the North Carolina State Board of Education were based on the median scaled-score standards recommended by all participants for the NTE General Knowledge and Communication Skills Tests, each reduced by two total standard errors, computed in the scaled-score metric. Thus the second, more stringent, error model was used to develop final recommended standards.

Distributions of standards recommended by each type of participant, with and without adjustments for errors of measurement and estimation, were computed from recommendations obtained during each standard-setting session. Salient statistics on these distributions are shown in both tabular and graphical form in Section III of this report.

Analysis of Participant's Comments on Tests and Procedures

Each subtest review group had an assistant who worked with the group leader in distributing materials and answering questions. A primary responsibility of the assistant was to act as an observer during the data-gathering sessions. The assistant recorded any comments or questions from the participants. In addition, participants were asked to record their comments and questions on the evaluaton forms provided. The comments recorded on the evaluation forms are reported verbatim in Appendix C . A staff member reviewed the comments and questions and categorized these into (1) comments and questions regarding the validation and standard-setting procedures (2) comments and questions regarding the NTE tests and their intended uses and (3) an "other" category, containing general comments and questions. A narrative summary of the comments and questions made by participants during the data-gathering sessions follows.

a. Participant's Comments and Questions Regarding Validation and Standard-Setting Procedures

The issue most frequently commented on concerned evaluation of the subtests in terms of minimally competent students. Many participants felt there was a difference between the skills actually possessed by a minimally competent student and skills which entering teacher education students should have. That is, there was a tendency to confuse "ideal" skills with those that were described in the directions, i.e., those of a minimally competent student. It appears that participants understood that the judgements were to be made in terms of minimal competency. However, when requested to make judgements based on that conceptualization, some cognitive dissonance may have been created by a strong tendency to think of what "ought to be".

There was also a question raised by several participants as to whether all the participants used the same criteria for judging minimal competence. Some particiants felt that perhaps different types of participants (i.e., public school teachers and college and university faculty; education faculty and subject-area faculty; public university faculty and private university faculty; etc.) might differ in their judgement of what skills the minimally competent student should possess.

Several questions arose regarding the population of students which had taken the NTE in the spring of 1983 and on which all statistical information was based (i.e., item data and cumulative frequency data). Questions were mainly concerned with who was tested, whether or not they were volunteers, what schools they attended, what percent were minority, etc. In general, the participants wanted to know how representative the students tested in the Spring of 1983 are of the general population of applicants to teacher eduation programs. The sentiment was that more information would have been useful in making a valid interpretation of the statistical data and therefore would have helped in setting standards.

Some participants had questions as to how the statistical information was calculated and how it was to be used in setting standards. In several instances, the principal investigators were called in to answer these questions. The observers' notes suggest that the questions were answered to the satisfaction of the participants and that the participants had a sufficient understanding of the statistical information to make valid use of it.

Several participants commented that providing the statistics based on the Spring 1983 testing helped give a good context for setting standards. However, there were those who commented that they felt they were being coerced into changing the standards they had originally set because these standards would result in either too many or too few students passing the test. A small number of participants felt that perhaps the standard had actually been pre-determined and that they were being coerced into legitimizing this pre-determined standard.

Some participants had questions as to their own ability to judge the validity of the tests they were reviewing. For example, an English department faculty member reviewing the reading test stated that while she/he taught literature couses she/he did not teach reading per se and therefore questioned the validity of her/his judgements regarding a test of reading skills.

Some participants felt that it was difficult to assess instructional validity, since they did not know what students had actually been taught in high school and college classrooms. They pointed out that, although there were classes with certain course titles, they did not know what material was actually covered. Some also questioned whether or not some important skills, such as the drawing of inferences, can be taught. Several participants commented that, although the applicants may have been exposed to certain skills they may not have had any opportunity to practice the skills and therefore become proficient in their use. However, most participants expressed confidence in their ability to judge the validity of the tests they were asked to review.

Most participants voiced satisfaction with the overall process of validation and standard setting and supported the need for a screening instrument. However, one participant did refuse to take part in the validity and standard setting, stating that he/she had "serious reservations" about the entire process and was "concerned about what we are doing to our students." There were several other participants who felt that an objective test such as the NTE could not be used as a valid instrument for determining success in teacher education programs. The need for information concerning the predictive validity of the NTE for predicting success in North Carolina teacher education programs was mentioned.

b. Participant Comments and Questions Regarding the NTE Tests and Their Entended Uses

Most participants commented favorably on the overall validity of the subtests they reviewed. However, most commented that certain items were ambiguous and should be rewritten. Particularly on the Science subtest, many participants felt that there were several items in which the wording was poor and, in some cases, inaccurate. Participants also expressed the feeling that some items may have measured different skills or knowledge than they were intended to measure. Several participants commented that having more information on what each item was attempting to measure would have been useful.

Many of the participants reviewing the Literature and Fine Arts subtest commented that the subtest appeared to concentrate mainly on certain areas rather than tapping the general area of literature and fine arts. In particular, it was felt that art was over-emphasized and that music and literature were under-emphasized. A problem for the participants on the Literature and Fine Arts subtest was the poor quality reproduction. Most participants commented that the poor quality of the photographs on the test made answering some items virtually impossible. The Science and Literature and Fine Arts subtests were by far the most heavily criticized, based on the participant's comments.

A few participants questioned the use of an objective test for measuring the skills and knowledge the NTE is intended to measure. They expressed the feeling that the NTE is more a measure of test-taking skills and problem solving ability than it is a measure of specific knowledge or skill areas.

Some participants commented that a test should be developed specifically for the purpose of screening applicants to teacher education programs in North Carolina rather than using an instrument already developed for another use. There also was a question as to whether or not only one test should be used to screen applicants to teacher education programs rather than several different sources of data.

There were numerous comments and questions concerning the logistics of using the NTE as a screening instrument. Several participants questioned the need to use NTE scores rather than scores on the SAT or other tests already in use. One participant brought with him/her information showing a high correlation between SAT and NTE scores among students taking both tests at his/her institution. He/she stated that using the NTE in addition to the SAT resulted in a duplication of information, extra cost for the student, and more paper work for the teacher education program.

Other logistical questions raised by the participants involved when and where the test would be administered, how the testing program would be funded, whether or not applicants would be allowed to re-take the test, and whether or not out-of-state students would be required to take the NTE. There was also a question raised concerning whether the major use of the test results would be to keep students out of teacher education programs, to predict success in programs, or to determine who needs remediation.

c. Other Participant Comments

Comments regarding the facilities in Asheville and Graham were generally positive. However, many participants noted problems with the physical arrangements at the Goldsboro data-collection site. The facilty used in Goldsboro was the Officer's Club at Seymour Johnson Air Force base. Most groups were divided only by partitions and many participants commented on the problem of noise, which caused difficulties in hearing the group leader and in hearing other participants during discussion periods. A number of participants also commented on the poor lighting. This was a particular problem for the Social Studies and Writing-Essay groups,that met in the lounge area of the club.

Analysis of participants' evaluation data. The final element of the data-collection procedure was to have participants evaluate the activities of validation and standard setting. A single evaluation questionnaire consisting of 25 multiple-choice items was administered to all participants. Responses to evaluation items were recorded on computer-readable scoring sheets which also contained space for additional comments. Response choices for each item were: SD--Strongly Disagree, D--Disagree, DK--Don't Know, A--Agree, SA--Strongly Agree.

All participants were asked to respond to every item. Those for whom certain items were not relevant (e.g., public school teachers did not participate in the instructional validity session) were asked to record "Don't Know" as their response. The logical consistency of participants' responses was checked, and where responses were incorrectly marked (e.g. a public school teacher recording a response other than "Don't Know" to an instructional validity item) responses were corrected and percentages were adjusted accordingly.

For purposes of analysis, items on the Participant's Evaluation Questionnaire were grouped under seven headings that reflect the topics addressed by the questionnaire. The tabular results of these analyses are reported in Tables 3-9. Each table contains an item number, a concise statement of the question asked, a listing of each response option, from "Omit" to "Strongly Agree", the frequency of participants' responses to each option, the percent of participants who responded to each option, and the cumulative percent of responses to each option.

The first series of questions to be analyzed pertained to participants' perceptions ofthe need for the North Carolina Teacher Quality Assurance Program (Table 3). The items included in this grouping were numbered 4, 5 and 6 on the Questionnaire. Item 4 asked about participants' perceptions of the need to screen candidates for teacher education programs. 90% of the participants disagreed with the statement that "it is not important to screen candidates." Item 5 asked whether the participants thought that standardized testing was an appropriate way to screen candidates for teacher education programs. Almost 60% of the participants agreed with the statement, 23% disagreed, and 17% didn't know. Item 6 asked participants to respond to the statement that the NTE tests were not appropriate for screening applicants to teacher education programs. 58% of the participants disagreed with the statement, 25% agreed with it, and 16% indicated that they did not know. In summary, a majority of the participants strongly support a main objective of the North Carolina Teacher Quality Assurance Program, that of screening candidates to teacher education programs. Also, a majority of the participants agreed that standardized tests are appropriate for this purpose, and that the NTE tests, in particular, are appropriate devices.

The second series of questions to be analyzed pertained to participants' perceptions of the need for validiation and standard setting for the NTE. Analyses of the responses to these items are reported in Table 4. The items included in this grouping were numbered 1, 2 and 3. Item 1 asked about

TABLE 3
PARTICIPANTS' PERCEPTIONS OF THE NEED FOR THE NORTH CAROLINA TEACHER QUALITY ASSURANCE PROGRAM

```
ITEM4     NOT IMPORTANT TO SCREEN CANDIDATES

                                                             VALID      CUM
 VALUE LABEL                   VALUE  FREQUENCY  PERCENT   PERCENT   PERCENT

OMIT                               0          1       .3        .3        .3
STRONGLY DISAGREE                  1        199     65.9      65.9      66.2
DISAGREE                           2         72     23.8      23.8      90.1
DON'T KNOW                         3         11      3.6       3.6      93.7
AGREE                              4          8      2.6       2.6      96.4
STRONGLY AGREE                     5         11      3.6       3.6     100.0
                                          -------  -------   -------
                               TOTAL        302    100.0     100.0

VALID CASES     302      MISSING CASES     0
- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -

ITEM5     STANDARDIZED TESTING APPROPRIATE SCREEN

                                                             VALID      CUM
 VALUE LABEL                   VALUE  FREQUENCY  PERCENT   PERCENT   PERCENT

OMIT                               0          2       .7        .7        .7
STRONGLY DISAGREE                  1         25      8.3       8.3       8.9
DISAGREE                           2         43     14.2      14.2      23.2
DON'T KNOW                         3         52     17.2      17.2      40.4
AGREE                              4        144     47.7      47.7      88.1
STRONGLY AGREE                     5         36     11.9      11.9     100.0
                                          -------  -------   -------
                               TOTAL        302    100.0     100.0

VALID CASES     302      MISSING CASES     0
-----------------------------------------------------------------------------
ITEM6     NTE TESTS NOT APPROPRIATE SCREEN

                                                             VALID      CUM
 VALUE LABEL                   VALUE  FREQUENCY  PERCENT   PERCENT   PERCENT

OMIT                               0          3      1.0       1.0       1.0
STRONGLY DISAGREE                  1         53     17.5      17.5      18.5
DISAGREE                           2        121     40.1      40.1      58.6
DON'T KNOW                         3         48     15.9      15.9      74.5
AGREE                              4         50     16.6      16.6      91.1
STRONGLY AGREE                     5         27      8.9       8.9     100.0
                                          -------  -------   -------
                               TOTAL        302    100.0     100.0

VALID CASES     302      MISSING CASES     0
```

TABLE 4
PARTICIPANTS' PERCEPTIONS OF THE NEED FOR VALIDATION AND STANDARD SETTING FOR THE NTE

ITEM1   UNDERSTAND NEED TO VALIDATE THE NTE

VALUE LABEL	VALUE	FREQUENCY	PERCENT	VALID PERCENT	CUM PERCENT
OMIT	0	1	.3	.3	.3
DISAGREE	2	6	2.0	2.0	2.3
DON'T KNOW	3	1	.3	.3	2.6
AGREE	4	100	33.1	33.1	35.8
STRONGLY AGREE	5	194	64.2	64.2	100.0
	TOTAL	302	100.0	100.0	

VALID CASES   302   MISSING CASES   0

- -

ITEM2   DID NOT UNDERSTAND NEED TO SET STANDARDS

VALUE LABEL	VALUE	FREQUENCY	PERCENT	VALID PERCENT	CUM PERCENT
OMIT	0	3	1.0	1.0	1.0
STRONGLY DISAGREE	1	182	60.3	60.3	61.3
DISAGREE	2	85	28.1	28.1	89.4
DON'T KNOW	3	4	1.3	1.3	90.7
AGREE	4	24	7.9	7.9	98.7
STRONGLY AGREE	5	4	1.3	1.3	100.0
	TOTAL	302	100.0	100.0	

VALID CASES   302   MISSING CASES   0

ITEM3   IMPORTANT TO KNOW RELEVANCE

VALUE LABEL	VALUE	FREQUENCY	PERCENT	VALID PERCENT	CUM PERCENT
OMIT	0	2	.7	.7	.7
STRONGLY DISAGREE	1	4	1.3	1.3	2.0
DISAGREE	2	4	1.3	1.3	3.3
DON'T KNOW	3	4	1.3	1.3	4.6
AGREE	4	96	31.8	31.8	36.4
STRONGLY AGREE	5	192	63.6	63.6	100.0
	TOTAL	302	100.0	100.0	

VALID CASES   302   MISSING CASES   0 participants' perceptions of the need to validate the NTE for use as a screening device for teacher education programs. 97% of the participants agreed with the need to validate the NTE for this purpose. Item 2 asked participants to respond to the statement that they "did not understand the need to set standards." 88% of the participants disagreed with this statement. Item 3 asked participants whether it was important to know the relevance of the NTE to success in teacher education programs. More than 95% of the participants agreed with the need to know the relevance of the NTE. In summary, a majority of the participants strongly support the purpose of the study, that is to validate the NTE for use as a screening device for teacher education programs, and to set standards of competence for individuals taking the test.

The third series of questions to be analyzed examined participants' perceptions of the adequacy and clarity of the validation procedures. Items from the evaluation questionnaire included in Table 5 were numbered 7, 8, 11 and 24. Item 7 asked if the participants had time to complete the NTE test. Only participants making judgments about the seven multiple-choice tests were to comment about this item. The 61 participants judging the Writing essay test were directed to mark "Don't Know" on the evaluation form. Two analyses were performed for this question. Item 7a in Table 5 shows percents for all participants, including the 61 Essay participants. Item 7b reports only the responses of participants who took a NTE multiple-choice test. 59% of those taking a NTE multiple-choice test agreed that they did have time to finish, 23% did not know if they had time to finish, and almost 19% disagreed that they

TABLE 5
PARTICIPANTS' PERCEPTIONS OF THE ADEQUACY AND CLARITY OF THE VALIDATION PROCEDURES

```
ITEM7a     HAD TIME TO COMPLETE NTE TEST

                                                               VALID      CUM
  VALUE LABEL                 VALUE  FREQUENCY   PERCENT    PERCENT   PERCENT

STRONGLY DISAGREE                 1         18       6.0        6.0       6.0
DISAGREE                          2         38      12.6       12.6      18.6
DON'T KNOW                        3         68      22.5       22.5      41.1
AGREE                             4         92      30.5       30.5      71.5
STRONGLY AGREE                    5         86      28.5       28.5     100.0
                                       -------   -------    -------
                              TOTAL        302     100.0      100.0

VALID CASES      302       MISSING CASES      0
- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -----
ITEM7b     HAD TIME TO COMPLETE NTE TEST


                                                               VALID      CUM
  VALUE LABEL                 VALUE  FREQUENCY   PERCENT    PERCENT   PERCENT

STRONGLY DISAGREE                 1         18       7.0        7.0       7.0
DISAGREE                          2         38      16.0       16.0      23.0
DON'T KNOW                        3          7       3.0        3.0      26.0
AGREE                             4         92      38.0       38.0      64.0
STRONGLY AGREE                    5         86      36.0       36.0     100.0
                                       -------   -------    -------
                              TOTAL        241     100.0      100.0

VALID CASES      241       MISSING CASES      0

- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - ----
ITEM8     DID NOT UNDERSTAND RELEVANCE DIRECTIONS

                                                               VALID      CUM
  VALUE LABEL                 VALUE  FREQUENCY   PERCENT    PERCENT   PERCENT

OMIT                              0          1        .3         .3        .3
STRONGLY DISAGREE                 1         91      30.1       30.1      30.5
DISAGREE                          2        158      52.3       52.3      82.8
DON'T KNOW                        3          4       1.3        1.3      84.1
AGREE                             4         41      13.6       13.6      97.7
STRONGLY AGREE                    5          7       2.3        2.3     100.0
                                       -------   -------    -------
                              TOTAL        302     100.0      100.0

VALID CASES      302       MISSING CASES      0
```

TABLE 5 CONTINUED

```
ITEM11    HAD TIME TO MAKE RELEVANCE JUDGMENTS

                                                              VALID     CUM
 VALUE LABEL                  VALUE  FREQUENCY   PERCENT   PERCENT  PERCENT

OMIT                              0          1        .3        .3       .3
STRONGLY DISAGREE                 1          7       2.3       2.3      2.6
DISAGREE                          2         22       7.3       7.3      9.9
DON'T KNOW                        3          7       2.3       2.3     12.3
AGREE                             4        177      58.6      58.6     70.9
STRONGLY AGREE                    5         88      29.1      29.1    100.0
                                     -------   -------   -------
                              TOTAL        302     100.0     100.0

VALID CASES      302       MISSING CASES      0
-------------------------------------------------------------------------
ITEM24a     INSTRUCTIONAL VALIDITY  PROCEDURES CLEAR

                                                              VALID     CUM
 VALUE LABEL                  VALUE  FREQUENCY   PERCENT   PERCENT  PERCENT

OMIT                              0          3       1.0       1.0      1.0
STRONGLY DISAGREE                 1          2        .7        .7      1.7
DISAGREE                          2         14       4.8       4.8      6.5
DON'T KNOW                        3        104      33.4      33.4     39.9
AGREE                             4        116      39.2      39.2     79.0
STRONGLY AGREE                    5         63      21.0      21.0    100.0
                                     -------   -------   -------
                              TOTAL        302     100.0     100.0

VALID CASES      302       MISSING CASES      0

---------------------------------------------------------------------------
ITEM24b     INSTRUCTIONAL VALIDITY  PROCEDURES CLEAR

                                                              VALID     CUM
 VALUE LABEL                  VALUE  FREQUENCY   PERCENT   PERCENT  PERCENT

OMIT                              0          3       1.5       1.5      1.5
STRONGLY DISAGREE                 1          2       1.0       1.0      2.5
DISAGREE                          2         14       6.9       6.9      9.3
DON'T KNOW                        3          9       4.4       4.4     13.7
AGREE                             4        116      56.9      56.9     70.6
STRONGLY AGREE                    5         60      29.4      29.4    100.0
                                     -------   -------   -------
                              TOTAL        204     100.0     100.0

VALID CASES      204       MISSING CASES      0
```

had time to finish. Item 8 asked if the participants did not understand the relevance directions. 82% indicated that they disagreed with the statement that they did not understand the directions, 16% agreed with the statement. Item 11 asked if participants had time to make relevance judgments. 88% of participants agreed that they had time for making relevance judgments. Item 24 asked if the instructional validity procedures were clear. This question was directed to the college and university faculty participants only, public school participants were instructed to respond "Don't Know" to this question. Separate analyses were done for both the total number of participants (Item 24a) and for college and university faculty participants only (Item 24b). Of those participants who made judgments about instructional validity, over 86% agreed tnat the instructional validity procedures were clear. About 8% disagreed, and about 4% did not know.

To summarize the results shown in Table 5, a majority of participants (84%) had time to complete the test; disagreed with the statement that they did not understand the relevance directions; strongly agreed that they had time to make relevance judgments; and strongly agreed that the instructional validity instructions were clear.

The fourth series of questions to be analyzed examined participants' confidence in the validity judgments they made. The results of these analyses are shown in Table 6. Items on the evaluation questionnaire pertaining to this area were numbered 9, 10, and 25. Item 9 asked the particpants to express their opinion on whether it was difficult to make relevance judgments.

TABLE 6
PARTICIPANTS' PERCEPTIONS OF CONFIDENCE IN VALIDITY JUDGMENTS

```
ITEM9   DIFFICULT TO MAKE RELEVANCE JUDGMENTS

                                                          VALID      CUM
  VALUE LABEL                VALUE  FREQUENCY  PERCENT  PERCENT  PERCENT

OMIT                             0          1       .3       .3       .3
STRONGLY DISAGREE                1         31     10.3     10.3     10.6
DISAGREE                         2        131     43.4     43.4     54.0
DON'T KNOW                       3          8      2.6      2.6     56.6
AGREE                            4        106     35.1     35.1     91.7
STRONGLY AGREE                   5         25      8.3      8.3    100.0
                                      -------  -------  -------
                             TOTAL        302    100.0    100.0

VALID CASES      302       MISSING CASES      0


- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -


ITEM10    CONFIDENT ABOUT RELEVANCE JUDGMENTS

                                                          VALID      CUM
  VALUE LABEL                VALUE  FREQUENCY  PERCENT  PERCENT  PERCENT

OMIT                             0          2       .7       .7       .7
STRONGLY DISAGREE                1          7      2.3      2.3      3.0
DISAGREE                         2         46     15.2     15.2     18.2
DON'T KNOW                       3         32     10.6     10.6     28.8
AGREE                            4        167     55.3     55.3     84.1
STRONGLY AGREE                   5         48     15.9     15.9    100.0
                                      -------  -------  -------
                             TOTAL        302    100.0    100.0

VALID CASES      302       MISSING CASES      0
```

TABLE 6 CONTINUED

```
ITEM25a    CONFIDENT ABOUT INSTRUCTIONAL VALIDITY RATINGS

                                                            VALID     CUM
 VALUE LABEL                  VALUE  FREQUENCY   PERCENT   PERCENT  PERCENT

OMIT                              0          4       1.3       1.3      1.3
STRONGLY DISAGREE                 1          5       1.7       1.7      3.0
DISAGREE                          2         24       7.9       7.9     10.9
DON'T KNOW                        3        122      39.6      39.6     50.4
AGREE                             4        101      33.4      33.4     83.9
STRONGLY AGREE                    5         46      15.9      15.9    100.0
                                       -------   -------   -------
                              TOTAL        302     100.0     100.0

VALID CASES     302      MISSING CASES     0
```

```
ITEM25b    CONFIDENT ABOUT INSTRUCTIONAL VALIDITY RATINGS

                                                            VALID     CUM
 VALUE LABEL                  VALUE  FREQUENCY   PERCENT   PERCENT  PERCENT

OMIT                              0          4       2.0       2.0      2.0
STRONGLY DISAGREE                 1          5       2.5       2.5      2.5
DISAGREE                          2         24      11.8      11.9     11.8
DON'T KNOW                        3         24      11.8      11.8     27.9
AGREE                             4        101      49.5      49.5     77.5
STRONGLY AGREE                    5         46      22.5      22.5    100.0
                                       -------   -------   -------
                              TOTAL        204     100.0     100.0

VALID CASES     204      MISSING CASES     0
```

About 54% disagreed that these judgments were difficult to make, while 43% agreed that they were difficult judgments to make. Item 10 asked whether the participants were confident about their relevance judgments. Over 71% agreed that they were confident about the judgments, while over 17% disagreed about feeling confident of their judgments. Item 25 asked college and university participants whether they were confident about the instructional validity ratings they set. Public school teachers did not participate in the instructional validity sessions, and were asked to mark "Don't Know" for Item 25 on their evaluation forms. Two analyses of these data were completed; one for the total group of participants (Item 25a) and another for college and university faculty members only (Item 25b). Data from Item 25b indicate that about 73% of the participants in the instructional validity sessions agreed that they were confident about the instructional validity ratings they set, while about 15% stated that they disagreed. To summarize Table 6, the participants generally felt confident about their relevance and instructional validity judgments. Fifty-four percent disagreed that the relevance judgments were difficult to make; 71% agreed that they were confident about the relevance judgments they made; and, 73% of college and university participants agreed that they were confident about the instructional validity judgments they made.

The fifth series of questions analyzed examined participants' perceptions of the adequacy and clarity of the standard-setting procedures. Analyses for this series of questions are reported in Table 7. Items from the evaluation questionnaire that are included in this table were numbered 12, 15, 16 and 17. Items 12, 15 and 16 were directed to participants in groups that reviewed the multiple-choice tests. Participants in the Essay Writing Test group were instructed to mark "Don't Know" for Items 12, 15, and 16. Two analyses were completed for each of these items. Analyses numbered 12a, 15a and 16a show results for all 302 participants. Analyses numbered 12b, 15b and 16b show the results for the multiple-choice test participants only. These analyses will be discussed. Item 12 asked participants whether they understood the probability scale used in setting standards. 94% of the participants agreed that they understood the probability scale. Item 15 asked if the participants did not understand the percents in parentheses. About 97% of the participants disagreed that they did not understand the percents in parentheses. Item 16 asked participants whether they understood the numbers in boxes. Almost 95% agreed that they did understand the numbers in boxes used in the standard-setting sessions. Finally, Item 17 asked if the graph used in the standard-setting sessions made sense. Over 94% of the participants agreed that the graph used for the standard-setting sessions made sense. In summary, the participants responded strongly in support of the adequacy and clarity of the standard-setting procedures.

The sixth series of questions on the evaluation questionnaire pertained to participants' perceptions of confidence in their standard setting judgments and of their behavior in the standard-setting sessions. Items in this series

TABLE 7
PARTICIPANTS' PERCEPTIONS OF ADEQUACY AND CLARITY OF STANDARD SETTING PROCEDURES

```
ITEM12a    UNDERSTAND PROBABILITY SCALE

                                                             VALID      CUM
 VALUE LABEL                  VALUE  FREQUENCY  PERCENT   PERCENT   PERCENT

OMIT                              0          1       .3        .3        .3
STRONGLY DISAGREE                 1          1       .3        .3        .6
DISAGREE                          2          6      2.0       2.0       2.6
DON'T KNOW                        3         67     22.2      22.2      24.8
AGREE                             4        144     47.7      47.7      72.5
STRONGLY AGREE                    5         83     27.5      27.5     100.0
                                        -------  -------   -------
                              TOTAL        302    100.0     100.0

VALID CASES     302      MISSING CASES      0
```

- -

```
ITEM12b    UNDERSTAND PROBABILITY SCALE

                                                             VALID      CUM
 VALUE LABEL                  VALUE  FREQUENCY  PERCENT   PERCENT   PERCENT

STRONGLY DISAGREE                 1          1       .4        .4        .4
DISAGREE                          2          6      2.5       2.5       2.5
DON'T KNOW                        3          7      2.9       2.9       5.8
AGREE                             4        144     59.8      59.8      65.6
STRONGLY AGREE                    5         83     34.4      34.4     100.0
                                        -------  -------   -------
                              TOTAL        241    100.0     100.0

VALID CASES     241      MISSING CASES      0
```

TABLE 7 CONTINUED

```
ITEM15a    DID NOT UNDERSTAND PERCENTS IN PARENTHESES


                                                           VALID      CUM
 VALUE LABEL                VALUE  FREQUENCY  PERCENT    PERCENT   PERCENT

STRONGLY DISAGREE               1        119     39.4       39.4      39.4
DISAGREE                        2        114     37.7       37.7      77.2
DON'T KNOW                      3         62     20.5       20.5      97.7
AGREE                           4          5      1.7        1.7      99.3
STRONGLY AGREE                  5          2       .7         .7     100.0
                                     -------  -------    -------
                            TOTAL        302    100.0      100.0

VALID CASES     302      MISSING CASES     0

----------------------------------------------------------------------
```

```
ITEM15b    DID NOT UNDERSTAND PERCENTS IN PARENTHESES


                                                           VALID      CUM
 VALUE LABEL                VALUE  FREQUENCY  PERCENT    PERCENT   PERCENT

STRONGLY DISAGREE               1        119     49.3       49.3      49.3
DISAGREE                        2        114     47.3       47.3      96.6
DON'T KNOW                      3          1       .4         .4      97.1
AGREE                           4          5      2.0        2.0      99.1
STRONGLY AGREE                  5          2       .8         .8     100.0
                                     -------  -------    -------
                            TOTAL        241    100.0      100.0

VALID CASES     241     MISSING CASES 0
```

TABLE 7 CONTINUED

```
ITEM16a    UNDERSTOOD NUMBERS IN BOXES

                                                     VALID      CUM
 VALUE LABEL               VALUE  FREQUENCY  PERCENT  PERCENT  PERCENT

STRONGLY DISAGREE              1          2       .7       .7       .7
DISAGREE                       2          9      3.0      3.0      3.7
DON'T KNOW                     3         63     20.8     20.8     24.5
AGREE                          4        116     38.4     38.4     62.9
STRONGLY AGREE                 5        112     37.1     37.1    100.0
                                    -------  -------  -------
                           TOTAL        302    100.0    100.0

VALID CASES     302      MISSING CASES     0

------------------------------------------------------------------------
ITEM16b    UNDERSTOOD NUMBERS IN BOXES


                                                     VALID      CUM
 VALUE LABEL               VALUE  FREQUENCY  PERCENT  PERCENT  PERCENT

STRONGLY DISAGREE              1          2       .8       .8       .8
DISAGREE                       2          9      3.7      3.7      3.7
DON'T KNOW                     3          2       .8       .8      5.3
AGREE                          4        116     48.1     48.1     53.4
STRONGLY AGREE                 5        112     46.5     46.5    100.0
                                    -------  -------  -------
                           TOTAL        241    100.0    100.0

VALID CASES     241      MISSING CASES     0
------------------------------------------------------------------------

ITEM17    GRAPH FOR STANDARDS MADE SENSE

                                                     VALID      CUM
 VALUE LABEL               VALUE  FREQUENCY  PERCENT  PERCENT  PERCENT

OMIT                           0          1       .3       .3       .3
STRONGLY DISAGREE              1          8      2.6      2.6      3.0
DISAGREE                       2          7      2.3      2.3      5.3
DON'T KNOW                     3          1       .3       .3      5.6
AGREE                          4        153     50.7     50.7     56.3
STRONGLY AGREE                 5        132     43.7     43.7    100.0
                                    -------  -------  -------
                           TOTAL        302    100.0    100.0

VALID CASES     302      MISSING CASES     0
```

were numbered 13, 18 and 19. Analyses of these items are reported in Table 8. Items 13, 18, and 19 were intended for those participants who made standard-setting judgments about multiple-choice subtests of the NTE. Participants who reviewed the Essay Writng Test were asked to mark "Don't Know" for these items. Each of these items has been analyzed twice, once for all participants, (Items 13a, 18a, and 19a) and then only for participants who recommend standards for the multiple-choice tests (Items 13b, 18b, and 19b). The discussion will be restricted to analyses of Items 13b, 18b, and 19b. Item 13 asked participants if they were not confident about their probability estimates. About 59% of the participants disagreed that they were not confident about probability estimates, while about 35% agreed that they were not confident. Item 18 asked whether, in the iterative stage of the standard-setting process, participants had changed their estimated probabilities. Over 64% agreed that they had changed probabilities, and about 34% disagreed with the statement. Item 19 asked participants whether the estimated probabilities they set were lower during second session than they were during the first. Almost 39% agreed that they estimated lower probabilities during the second session than during the first, while 51% disagreed with the statement. In summary, while a majority of the participants (59%) agreed that they were confident about the probabilities they estimated, a large minority (35%) did not feel confident. A majority of the participants (64%) changed their probabilities during the iterative standard-setting procedure, and a small majority (51%)

TABLE 8
PARTICIPANTS' PERCEPTIONS OF CONFIDENCE IN STANDARD SETTING JUDGMENTS AND OF THEIR BEHAVIOR

```
ITEM13a   NOT CONFIDENT ABOUT PROBABILITY ESTIMATES

                                                     VALID      CUM
 VALUE LABEL             VALUE  FREQUENCY  PERCENT  PERCENT  PERCENT

STRONGLY DISAGREE            1         29      9.6      9.6      9.6
DISAGREE                     2        113     37.6     37.6     49.2
DON'T KNOW                   3         75     24.6     24.6     71.8
AGREE                        4         73     24.2     24.2     96.0
STRONGLY AGREE               5         12      4.0      4.0    100.0
                                  -------  -------  -------
                         TOTAL        302    100.0    100.0

VALID CASES     302      MISSING CASES      0
```

- -

```
ITEM13b   NOT CONFIDENT ABOUT PROBABILITY ESTIMATES

                                                     VALID      CUM
 VALUE LABEL             VALUE  FREQUENCY  PERCENT  PERCENT  PERCENT

STRONGLY DISAGREE            1         29     12.0     12.0     12.0
DISAGREE                     2        113     46.9     46.9     58.9
DON'T KNOW                   3         14      5.8      5.8     64.7
AGREE                        4         73     30.3     30.3     95.0
STRONGLY AGREE               5         12      5.0      5.0    100.0
                                  -------  -------  -------
                         TOTAL        241    100.0    100.0

VALID CASES     241      MISSING CASES      0
```

TABLE 8 CONTINUED

```
ITEM18a    DID NOT CHANGE ESTIMATED  PROBABILITIES

                                                        VALID     CUM
  VALUE LABEL                  VALUE  FREQUENCY  PERCENT  PERCENT  PERCENT

STRONGLY DISAGREE                  1         18      6.0      6.0      6.0
DISAGREE                           2         62     20.6     20.6     26.0
DON'T KNOW                         3         67     21.9     21.9     46.1
AGREE                              4        124     41.5     41.5     89.6
STRONGLY AGREE                     5         31     10.4     10.4    100.0
                                      -------  -------  -------
                               TOTAL        302    100.0    100.0

VALID CASES     302      MISSING CASES     0
```

- -

```
ITEM18b    DID NOT CHANGE ESTIMATED  PROBABILITIES

                                                        VALID     CUM
  VALUE LABEL                  VALUE  FREQUENCY  PERCENT  PERCENT  PERCENT

STRONGLY DISAGREE                  1         18      7.5      7.5      7.5
DISAGREE                           2         62     25.7     25.7     33.2
DON'T KNOW                         3          6      2.5      2.5     35.8
AGREE                              4        124     51.5     51.5     87.3
STRONGLY AGREE                     5         31     12.9     12.9    100.0
                                      -------  -------  -------
                               TOTAL        241    100.0    100.0

VALID CASES     241      MISSING CASES     0
```

TABLE 8 CONTINUED

```
ITEM19a    ESTIMATED PROBABILITIES LOWER SECOND SETTING THAN FIRST

                                                             VALID     CUM
 VALUE LABEL                 VALUE  FREQUENCY  PERCENT  PERCENT  PERCENT

OMIT                             0          3      1.0      1.0      1.0
STRONGLY DISAGREE                1         31     10.3     10.3     11.3
DISAGREE                         2         92     30.5     30.5     41.7
DON'T KNOW                       3         83     27.5     27.5     69.2
AGREE                            4         80     26.5     26.5     95.7
STRONGLY AGREE                   5         13      4.3      4.3    100.0
                                      -------  -------  -------
                             TOTAL        302    100.0    100.0

VALID CASES     302       MISSING CASES      0

------------------------------------------------------------------------


ITEM19b    ESTIMATED PROBABILITIES LOWER SECOND SETTING THAN FIRST

                                                             VALID     CUM
 VALUE LABEL                 VALUE  FREQUENCY  PERCENT  PERCENT  PERCENT

OMIT                             0          3      1.2      1.2      1.2
STRONGLY DISAGREE                1         31     12.9     12.9     14.1
DISAGREE                         2         92     38.2     38.2     52.3
DON'T KNOW                       3         22      9.1      9.1     61.4
AGREE                            4         80     33.2     33.2     94.6
STRONGLY AGREE                   5         13      5.4      5.4    100.0
                                      -------  -------  -------
                             TOTAL        241    100.0    100.0

VALID CASES     241       MISSING CASES      0
```

stated that they did not lower their probabilities on the second setting.

The final series of questions to be analyzed pertained to participants' perceptions of their use of data and discussion during the standard-setting procedures. The items included in this group of analyses were numbered 14, 20, 21, 22 and 23. Results are presented in Table 9. Items 14, 20, 21, and 23 were designed to be answered by participants involved in recommending standards for the multiple-choice tests only. Participants who reviewed the Essay Test were directed to mark "Don't Know" on their evaluation response forms. Data for each of these items have been analyzed twice, once for all participants (Items 14a, 20a, 21a, and 23a), and then only for participants who reviewed the multiple-choice subtests (Items 14b, 20b, 21b, and 23b). These latter analyses are discussed below. Item 14 asked participants whether they used the mini-test information. About 81% of the participants agreed that they used this information. Item 20 asked if data on the percent of all applicants had no influence on their decisions, 31% agreed that they had no influence, and about 64% disagreed with the statement. Item 21 asked participants whether data on the percent of applicants who were in the 30-50% range influenced their recommendations. Fifty-eight percent of the participants agreed that these data did influence them, and 36% disagreed with the statement. Item 22 asked participants whether the graph data had no influence on their recommendations. Forty-one percent of the participants disagreed that the graph data had no influence, while 53% agreed that it had no influence. Item 23 asked whether discussions held

TABLE 9
PARTICIPANTS' PERCEPTIONS OF THEIR USE OF DATA AND DISCUSSION IN STANDARD SETTING

```
ITEM14a    I USED THE MINI-TEST INFORMATION

                                                       VALID     CUM
 VALUE LABEL              VALUE  FREQUENCY  PERCENT  PERCENT  PERCENT

OMIT                          0          1       .3       .3       .3
STRONGLY DISAGREE             1          9      3.0      3.0      3.3
DISAGREE                      2         25      8.3      8.3     11.6
DON'T KNOW                    3         73     24.1     24.1     35.8
AGREE                         4        162     53.6     53.6     89.4
STRONGLY AGREE                5         32     10.6     10.6    100.0
                                   -------  -------  -------
                          TOTAL        302    100.0    100.0

VALID CASES     302       MISSING CASES     0
```

- -

```
ITEM14b    I USED THE MINI-TEST INFORMATION

                                                       VALID     CUM
 VALUE LABEL              VALUE  FREQUENCY  PERCENT  PERCENT  PERCENT

OMIT                          0          1       .4       .4       .4
STRONGLY DISAGREE             1          9      3.7      3.7      4.1
DISAGREE                      2         25     10.4     10.4     14.5
DON'T KNOW                    3         12      5.0      5.0     19.5
AGREE                         4        162     67.2     67.2     86.7
STRONGLY AGREE                5         32     13.3     13.3    100.0
                                   -------  -------  -------
                          TOTAL        241    100.0    100.0

VALID CASES     241       MISSING CASES     0
```

TABLE 9 CONTINUED

```
ITEM20a    DATA ON THE PERCENT OF ALL APPLICANTS HAD NO INFLUENCE

                                                           VALID     CUM
 VALUE LABEL                 VALUE  FREQUENCY  PERCENT   PERCENT  PERCENT

OMIT                             0          1       .3        .3       .3
STRONGLY DISAGREE                1         17      5.7       5.7      6.0
DISAGREE                         2        136     45.1      45.1     51.1
DON'T KNOW                       3         70     23.1      23.1     74.2
AGREE                            4         66     21.9      21.9     96.0
STRONGLY AGREE                   5         12      4.0       4.0    100.0
                                      -------  -------   -------
                             TOTAL        302    100.0     100.0

VALID CASES     302      MISSING CASES      0


-----------------------------------------------------------------------


ITEM20b    DATA ON THE PERCENT OF ALL APPLICANTS HAD NO INFLUENCE

                                                           VALID     CUM
 VALUE LABEL                 VALUE  FREQUENCY  PERCENT   PERCENT  PERCENT

OMIT                             0          1       .4        .4       .4
STRONGLY DISAGREE                1         17      7.1       7.1      7.5
DISAGREE                         2        136     56.4      56.4     63.9
DON'T KNOW                       3          9      3.7       3.7     67.6
AGREE                            4         66     27.4      27.4     95.0
STRONGLY AGREE                   5         12      4.0       4.0    100.0
                                      -------  -------   -------
                             TOTAL        241    100.0     100.0

VALID CASES     241      MISSING CASES      0
```

TABLE 9 CONTINUED

ITEM21a DATA ON THE PERCENT OF APPLICANTS IN 30-50% DID INFLUENCE

VALUE LABEL	VALUE	FREQUENCY	PERCENT	VALID PERCENT	CUM PERCENT
OMIT	0	1	.3	.3	.3
STRONGLY DISAGREE	1	19	6.3	6.3	6.6
DISAGREE	2	69	22.9	22.9	29.5
DON'T KNOW	3	73	24.1	24.1	53.6
AGREE	4	127	42.1	42.1	95.7
STRONGLY AGREE	5	13	4.3	4.3	100.0
	TOTAL	302	100.0	100.0	

VALID CASES 302 MISSING CASES 0

ITEM21b DATA ON THE PERCENT OF APPLICANTS IN 30-50% DID INFLUENCE

VALUE LABEL	VALUE	FREQUENCY	PERCENT	VALID PERCENT	CUM PERCENT
OMIT	0	1	.4	.4	.4
STRONGLY DISAGREE	1	19	7.9	7.9	8.3
DISAGREE	2	69	28.6	28.6	36.9
DON'T KNOW	3	12	5.0	5.0	41.9
AGREE	4	127	52.7	52.7	94.6
STRONGLY AGREE	5	13	5.4	5.4	100.0
	TOTAL	241	100.0	100.0	

VALID CASES 241 MISSING CASES 0

TABLE 9 CONTINUED

```
ITEM22    GRAPH DATA HAD NO INFLUENCE

                                                          VALID     CUM
  VALUE LABEL               VALUE  FREQUENCY  PERCENT   PERCENT   PERCENT

OMIT                            0          2       .7        .7        .7
STRONGLY DISAGREE               1         22      7.3       7.3       7.9
DISAGREE                        2        101     33.4      33.4      41.4
DON'T KNOW                      3         16      5.3       5.3      46.7
AGREE                           4        121     40.1      40.1      86.8
STRONGLY AGREE                  5         40     13.2      13.2     100.0
                                     -------  -------   -------
                            TOTAL        302    100.0     100.0

VALID CASES     302      MISSING CASES     0

- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
ITEM23a    DISCUSSION HAD INFLUENCE

                                                          VALID     CUM
  VALUE LABEL               VALUE  FREQUENCY  PERCENT   PERCENT   PERCENT

OMIT                            0          1       .3        .3        .3
STRONGLY DISAGREE               1         19      6.3       6.3       6.6
DISAGREE                        2         61     20.2      20.2      26.8
DON'T KNOW                      3         77     25.4      25.4      52.3
AGREE                           4        131     43.3      43.3      95.7
STRONGLY AGREE                  5         13      4.3       4.3     100.0
                                     -------  -------   -------
                            TOTAL        302    100.0     100.0

VALID CASES     302      MISSING CASES     0

------------------------------------------------------------------------
ITEM23b    DISCUSSION HAD INFLUENCE

                                                          VALID     CUM
  VALUE LABEL               VALUE  FREQUENCY  PERCENT   PERCENT   PERCENT

OMIT                            0          1       .4        .4        .4
STRONGLY DISAGREE               1         19      7.9       7.9       8.3
DISAGREE                        2         61     25.3      25.3      33.6
DON'T KNOW                      3         16      6.6       6.6      40.2
AGREE                           4        131     54.4      54.4      94.6
STRONGLY AGREE                  5         13      5.4       5.4     100.0
                                     -------  -------   -------
                            TOTAL        241    100.0     100.0

VALID CASES     241      MISSING CASES     0
```

during the standard-setting session had an influence on their recommendations. Sixty percent agreed that discussion did influence their recommendations, and 33% disagreed. In summary, a majority of participants were influenced in their standard-setting recommendations by these sources of information: 1) the mini-test information (81%), 2) data on the percent of all applicants tested in April, 1983 (64%), 3) data on the percent of applicants in the 30-50% range (58%), the graph data (53%), and discussions (60%).

In conclusion, analyses of responses to the Participants' Evaluation Questionnaire, indicate that the participants generally agreed with the need for the North Carolina Quality Assurance Program, and the need to validate and set standards for the NTE. In addition, a majority of the participants found that procedures for validation and standard-setting were clear and adequate. Similarly, a majority of participants felt confident about the judgments they made, and made use of information intended to help them in making recommendations when such information was available.

RESULTS

Participation in the Study

Overall participation. Six hundred eighty-two (682) persons were invited by letter to take part in the study. Each invitee was asked to reply and to indicate whether or not he/she would be able to participate. Those who responded affirmatively are referred to in this section of the report as "agreements." Those who actually took part in the study are referred to as "participants."

As shown in Table 10, 250,or 37% of the 682 invitees were public school teachers and the remaining 432 (63%) were college or university faculty members. Three hundred forty-seven (347) of 682 invitees agreed to participate. Of the 347 persons who agreed to participate, 105 (30%) were public school teachers and 242 (70%) were college or university faculty. Of the 306 actual participants, 100 (33%) were public school teachers and 206 (67%) were college or university faculty.

Table 11 shows the percent of public school teachers and college and university faculty who agreed to participate and the percent who actually participated. Of the 250 public school teachers invited, 105 agreed to participate and 100 actually participated. Thus the rate of agreement for public school teachers was 42% and the corresponding rate of participation was 40%. The rate of agreement and rate of participation among college and university faculty was somewhat higher than that of public school teachers. Of the 432 college and university invitees, 56% agreed to take part and 48% actually participated.

Table 10.

Number and percent of persons invited, those agreeing to participate, and those participating, by type of affiliation

	Invitees	Agreements	Participants
Public School Teachers	250 (37%)*	105 (30%)	100 (33%)
College and University Faculty	432 (63%)	242 (70%)	206 (67%)
Total	682	347	306

*Numbers in parentheses are percents of totals, by type of participant.

Table 11

Percent of persons agreeing to participate and actually participating, by type of affiliation.

	Agreements	Participants
Public School Teachers	42%	40%
College and Univesity Faculty	56%	48%
Total	51%	45%

Of the 682 invitees, 347 agreed to attend, resulting in an overall agreement rate of 51%. Three hundred-six of the 682 invitees actually took part in the study; the overall participation rate was 45%.

Participation by assignment to subtest. Table 12 presents information similar to that found in Tables 10 and 11, for each subtest review group. Information is presented for college and university faculty only. Data in Table 12 are limited to college faculty because public school teachers were assigned to tests only after they had given affirmative responses to letters of invitation. The reader should also note that persons who were invited to review the writing subtests were not assigned to the multiple-choice or essay review groups until they had agreed to participate. Therefore, percentages are reported only for combined writing panels (multiple-choice and essay).

As shown in Table 12, 68 college and university faculty members were invited to review the Listening subtest. Thirty-one (31) or 46% of those invited agreed to participate and 25 (37%) of those invited, actually participated. These percentages for the Listening subtest and those for the Mathematics subtest, are notably lower than for other subtests. Most of the agreement and participation rates for the other subtests are reasonably consistent with the overall percents of agreement (57%) and participation (48%) presented in Table 11. In contrast to the low percentages of agreement and participation in the Listening subtest group, those for the science subtest group were quite high. Of the 37 faculty members invited to review the Science subtest, 27 (73%), agreed to participate and 23 (62%), actually

Table 12

Number and percentage of college and university faculty agreeing to participate and actually participating, by assignment to subtest.

Communication Skills Subtests

	Listening	Writing (Multiple Choice and Essay)	Reading
Number Invited	68	154	38
Number and Percent of Invitees Who Agreed to Participate	31 (46%)*	89 (58%)	23 (61%)
Number and Percent of Invitees Who Actually Participated	25 (37%)	74 (48%)	22 (58%)

Table 12 (continued)

General Knowledge Subtests

	Science	Mathematics	Social Studies	Literature/ Fine Arts
Number Invited	37	38	44	53
Number and Percent of Invitees Who Agreed to Participate	27 (73%)	18 (47%)	24 (55%)	30 (57%)
Number and Percent of Invitees Who Actually Participated	23 (62%)	17 (45%)	23 (52%)	24 (45%)

* Percent of invitees, by subtest

participated. These percentages are well above the averages shown in Table 4.

Participation by faculty members from traditional minority and non-minority institutions. A special effort was made to ensure the participation of college and university faculty members from traditional minority institutions. Volume 3 of our Final Report for Phase II of this project (June 30, 1983, pp 24-34) outlines the procedures used to oversample faculty from traditional minority institutions. It should be noted that public school teachers who were invited to participate were randomly selected and randomly assigned to subtests and were not categorized as having attended a traditional minority or non-minority institution. Thirty-seven percent (37%) of college and university faculty members who participated in the study were from traditional minority institutions. Table 13 contains the number of participants in each subtest review group that were from traditional minority or traditional non-minority institutions. The percent of participants from traditional minority institutions ranged from 28% (Listening) to 45% (Reading). Twenty-five (25) college and university faculty members reviewed the Listening subtest. Of these, 7 (28%), were from traditional minority institutions and 18 (72%) were from traditional non-minority institutions. Therefore, the percent of college and university faculty who reviewed the Listening subtest who were from traditional minority institutions was somewhat lower than the overall percent of all college and university faculty who participated in the study. Numbers and percents of faculty who reviewed Writing multiple-choice and Writing essay

Table 13

Number and percent of faculty participants in each subtest review group from traditionally minority institutions non-minority institutions.

Communication Skills Subtests

	Listening	Writing Multiple choice	Writing Essay	Reading
Minority Institutionn	7 (28%)*	13 (35%)	13 (35%)	10 (45%)
Non-Minority Institutions	18 (72%)	24 (65%)	24 (65%)	12 (55%)
Total College/ University Faculty	25	37	37	22

Table 13 (continued).

General Knowledge Subtests

	Science	Mathematics	Social Studies	Literature/ Fine Arts
Minority Institutions	8 (35%)	7 (41%)	10 (43%)	9 (38%)
Non-Minority Institutions	15 (65%)	10 (59%)	13 (57%)	15 (62%)
Total College/ University Faculty	23	17	23	24

* Percent of total group, by subtest

subtests, who were from traditional minority institutions, were identical (13 or 35%) for both subtests, with 37 participants reviewing each test. The sizes of these review groups were also equal, and 24 faculty members, (65% of each group), were from traditional non-minority institutions. These percents, and those for the Science and Literature and Fine Arts subtests, are similar to the overall proportion of participants from minority institutions. Percents of participants from traditional minority institutions were slightly above the overall percent for the Reading, Mathematics and Social Studies subtest review groups. Table 14 contrasts traditional minority and non-minority institutions by reporting percents of invitees who agreed to participate and who actually participated. As shown in Table 14, on the Listening subtest panel, there were 22 invitees from traditional minority institutions. Of these, 8 (36%), agreed to participate and 7 (32%), actually participated. In contrast, 46 college and university faculty members from traditional non-minority institutions were invited to review the Listening subtest. Of these, 23 (50%) agreed to participate and 18 (39%), actually participated.

As was observed earlier, the percent of invitees who agreed to participate and the percent who actually participated were somewhat lower for the Listening subtest than for most of the other subtests. Reference to Table 14 reveals that the percent of faculty agreeing to participate and actually participated were lower among faculty members from traditional minority institutions than among those from non-minority institutions. Only 36% of the faculty from traditional minority institutions agreed to participate in the Listening Test panel.

Table 14

Breakdown of persons in each test panel from traditionally minority vs. traditionally non- minority institutions: number and percentage of invitees, those agreeing to participate, and those actually participating.

Communication Skills Subtests

Type of Institution	Listening Minority	Listening Non-Minority	Writing Multiple Choice and Essay Minority	Writing Multiple Choice and Essay Non-Minority	Reading Minority	Reading Non-Minority
Number of Invitees	22	46	40	114	18	20
Number and Percent of Invitees Agreeing to Participate	8 (36%)*	23 (50%)	29 (73%)	60 (53%)	10 (56%)	13 (65%)
Number and Percent of Invitees Actually Participating	7 (32%)	18 (39%)	26 (65%)	48 (42%)	10 (56%)	12 (60%)

Table 14 (continued)

General Knowledge Subtests

	Science		Mathematics		Social Studies		Literature and Fine Arts	
Type of Institution:	Minority	Non-Minority	Minority	Non-Minority	Minority	Non-Minority	Minority	Non-Minority
Number of Invitees	16	21	17	21	15	29	16	37
Number and Percentage of Invitees Agreeing to Participate	12(75%)	15(71%)	8(47%)	10(48%)	10(67%)	14(48%)	10(63%)	20(54%)
Number and Percentage of Invitees Actually Participating	8(50%)	15(71%)	7(41%)	10(48%)	10(67%)	13(45%)	9(56%)	15(41%)

* Percent of total group, by subtest

The numbers and percents in Table 14 for the Writing panel include both those faculty members reviewing the multiple- choice Writing subtest and those reviewing the Essay Writing subtest. As noted earlier, these participants were not assigned to groups until after they had agreed to participate. Therefore, it was not possible to separate invitees assigned to the Writing multiple-choice subtest from those assigned to the Essay subtest in Table 14. Of the total writing sample, 40 faculty invitees came from traditional minority institutions. Twenty-nine of these 40 (73%) agreed to participate and 26 (65%) actually participated. One hundred and fourteen faculty members from traditional non-minority institutions were invited to review the Writing subtests. Of these, 60 (53%), agreed to participate and 48 (42%), actually participated.

The percent of faculty members from traditionally minority institutions who agreed to participate was highest for the Science subtest, twelve of the 16 invitees, (75%), agreed to participate. However, only 8 of the 16 (50%) actually participated. Fifteen of the 21 (71%) faculty invitees from traditional non-minority institutions who were assigned to the Science subtest agreed to participate, representing the highest percent of agreement for faculty from non-minority institutions. In this case all 15 who agreed to participate, actually participated, so the rate of participation was also 71%. The highest rate of participation among faculty members from traditional minority institutions was on the Social Studies subtest. On this subtest, 10 of the 15 minority institution invitees (67%) agreed to participate and all 10 (67% of those invited) actually participated.

Appendix D1 reports the number of invitees by individual institution and by subtest. Institutions are identified in the rows of Appendix D1 and the subtest groups are listed in the columns. For example, two faculty members were invited from Barber-Scotia College to review the Listening subtest and two were invited to review the Writing subtest. Therefore, the total number of invitees from Barber-Scotia College, shown in the column marked Total, was 4. Three faculty members were invited from Bennett College to review the Mathematics subtest and three were invited to review the Writing subtest. Thus, the total number of invitees from Bennett College, listed in the right margin, was six. The largest number of invitees from any one institution was 48 from East Carolina University, followed by Appalachian State University with 36 and North Carolina Central University with 23. The number of invitees at a given insititution was based on the proportion of teacher education graduates from that institution. Detailed procedures are described in Volume 2 of the Final Report of Phase II of this project (June, 1983, pp 24-34).

The total number of invitees for each subtest is listed at the bottom of the table in Appendix D1. For example the total number of faculty invitees for the Science subtest review group was 27, and the total number of faculty invitees for the Mathematics subtest review group was 38. Although the size of the review group for the Writing subtest is the largest, it should be noted that this figure represents both the Writing

Multiple-Choice and the Writing Essay groups.

Appendix D2 shows the number of college and university faculty members who agreed to participate, by institution and by test. It is organized in the same way as the table in Appendix D1. Institutions are identified in the rows of the table in Appendix D2 and subtest panels are identified in the columns. For example, two faculty members from Barber-Scotia College agreed to review the Writing subtest. The two faculty invited to review the Listening subtest did not agree to participate. Thus, the total number of faculty members from Barber-Scotia College who agreed to participate, reported in the right margin, was two. The total number of faculty members from all colleges and universities who agreed to review the Listening subtest, listed in the bottom margin, was 31. The total number of college and university faculty members who agreed to participate across all subtests, was 242.

Appendix D3 reports the number of participants by institution and test. There were no participants from Barber-Scotia College, so the total number of faculty participants from Barber-Scotia, listed in the right margin, was zero. There was one participant from Bennett College who reviewed the Mathematics subtest and another who reviewed the Essay subtest. Therefore, the total number of faculty participants from Bennett College, was two. The largest number of participants from any one institution was 22, from East Carolina University. The total number of college and university faculty who reviewed each test is reported at the bottom of the table in Appendix D3. Twenty-five college and university faculty members reviewed the

Listening subtest, 37 reviewed the Writing multiple choice subtest, and so on. The number of public school teachers who reviewed each subtest is also listed at the bottom of Appendix D3. Eighteen public school teachers reviewed the Listening subtest, 14 reviewed the Writing Multiple-Choice subtest, etc. Finally, the grand total of college and university faculty and public school teachers who reviewed each subtest is listed in the bottom margin of the table in Appendix D3. A total of 43 participants reviewed the Listening subtest (25 college and university faculty members plus 18 public school teachers). A total of 206 college and university faculty members and 100 public school teachers participated in the study, for a grand total of 306 participants.

Test validity

Relevance of the tests

Communication Skills Test. The results of the relevance judgments for the Listening, Writing and Reading multiple-choice subtests are reported in Tables 15 to 17. The number of judgments at each of the four scale points are given for each item and for each subtest. The average and standard deviation on the four-point scale and for the total subtest are also presented. The percent of responses which were judged to be "crucial" or "important" is also given for each item and for each subtest. The individual responses to each of the content validity and relevance ratings for the Essay subtest are reported in Table 22. Table 23 provides aggregate data for subtests and summarizes the validity indices described an earlier section of this report. Table 32 evaluates and summarizes that evidence.

As a whole, the evidence describing the relevance of the skills and knowledge assessed by the Communication Skills Test appears to be consistent. This is especially true for the Reading and Writing multiple-choice subtests. Each of the five indices for both subtests was judged to be satisfactory. All of the items on both subtests had mean ratings equal to or greater than 2.25, ten percent or less of the responses were made in the "not relevant" category. More than 75% of the responses were made in the "crucial" or "important" categories. All items on both subtests were judged to be "crucial" or "important" by a majority of raters. Evidence on the Listening subtest is not quite as strong as that for the Reading and Writing subtests. Four of the five indices for the Listening subtest were satisfactory and one was just marginally satisfactory. The mean rating for the Listening subtest on the four-point scale was 3.03. All of the items on the test received ratings of 2.25 or greater and 10 percent of the judgments were rated "not relevant". All items were rated "crucial" or "important" by a majority of participants on the Listening subtest. A total of 74 percent of responses made on the four-point scale were made in the "crucial" or "important" categories, barely placing this index in the marginally satisfactory category.

The Essay subtest was evaluated by utilizing information from Questions #1 and #2. Results are reported in Table 22. The first question sought traditional content validity information. Does the NTE Essay Test measure general writing skills? A clear majority of the participants either "agreed" (37) or "strongly agreed" (7). While 10 participants were uncertain, only six "disagreed" or "strongly disagreed". Question number two is more directly related to the relevance issue. Participants were asked to rate the importance of the writing skills which were actually measured by the NTE Essay Test, to success in a teacher education program. Thirty-three participants rated the skills "crucial" and twenty-six rated them "important". No participants rated them "irrelevant".

In summary, it would appear that the skills and knowledge assessed by the NTE Communication Skills Test are judged to be important to success as a student in teacher education.

General Knowledge Test. The results of the relevance judgments for the Science, Social Studies, Literature/Fine Arts and Mathematics multiple- choice subtests are reported in Tables 18 to 21. The data are presented in the same format as that of Tables 15 to 17. Table 23 reports aggregate data for total subtests and summarizes the five validity indices for the multiple-choice subtests which were described in Section II. Table 32 evaluates and summarizes that evidence.

As a whole, evidence on the relevance of the General Knowledge Tests is a bit more mixed than is evidence for the Communication Skills Test. Indices for both the Literature/Fine Arts and Social Studies subtests are generally satisfactory. Four of the five indices were evaluated as satisfactory for both subtests. One index for the Literature/Fine Arts subtest was marginally satisfactory: only 71 percent of the responses marked were "crucial" or "important". One of the five indices for the Social Studies subtest was judged to be unsatisfactory: only 69 percent of responses on the Social Studies subtest were "crucial" or "important". The Mathematics subtest had two indices which were marginal but the other three were satisfactory. The satisfactory indices were: the average rating on the Mathematics subtest was 2.96, eighty-four percent of the items were rated "important" or "crucial" by a majority of raters, and only seven percent of the ratings were in the "not relevant" category.

The Science subtest relevance data are less favorable than that of any of the other subtests. Two of the indices were satisfactory, one was marginally satisfactory, and two were unsatisfactory. The two unsatisfactory sources of evidence were as follows: sixteen percent of the responses were in the "not relevant" category and only fifty-seven percent of the responses were to the "crucial" or "important" categories. On the positive side, 93% of the items received average ratings greater than 2.25 and the average for the whole test was 2.54. Relevance is clearly mixed for this subtest.

Because the ratings on the Science subtest are marginally satisfactory, we found it is useful to examine other sources of information which might amplify, or corroborate, the present observations. Table 24 provides an analysis of the relevance ratings by type of participant ie., college or university faculty and public school teachers. For each of the indices on the Science subtest there appears to be a consistent, if modest, difference between the two types of participants. University or college participants tended to regard the content of the Science subtest as being more important than did the public school teachers. (In general, there was a trend for the college/university participants to be more favorable in evaluating the relevance of the Science and Literature/Fine Arts subtests. The reverse was true for the Mathematics subtest.)

Was there any additional information available with which to evaluate the Science subtest? At the conclusion of data collection, each participant was asked to evaluate the activities of the day. Participants did this with the assistance of an evaluative questionnaire, the results for which are provided in another section of this report. However, one item on the questionnaire consisted of a Likert-style item which stated that the NTE was not an appropriate screening device. While almost 60% of all of the study participants disagreed with that statement, only 42% of the Science panel disagreed. This suggests that those who participated in the Science panel had less overall confidence in the NTE test which they reviewed than did all other participants. As assessed in another item, the difference between Science panelists' and all panelists' confidence in their relevance judgments was relatively small with Science panelists showing less confidence. Participants who expressed written comments offered a variety of observations which are summarized in Section II. Several participants observed that the subtest content and the state curriculum in Science did not seem consistent. One factor which might help explain the marginal ratings on the Science subtest is the relative complexity of the Scence subtest with respect to the number of differences between the disciplines measured by the subtest (ie., Biology, Botany, Physics, Earth Science, Chemistry etc.) If that hypothesis is correct, then one would expect General Knowledge Test ratings to be lowest for Science and Social Studies, the more complex of the four subtests, almost as low for the Literature/Fine Arts subtest and highest for the Mathematics subtest, which measures a single discipline.

In summary, we conclude that the Communication Skills Tests in Reading, Writing multiple-choice, Writing Essay and Listening all appear to be acceptable in terms of the relevance of what they assess, to successful performance in teacher education programs. The following General Knowledge subtests were found to be acceptable with regard to relevance: Mathematics, Literature/Fine Arts and Social Studies. The Science subtest was found to be only marginally acceptable. It is our judgment that continued study of the Science subtest be pursued by the State Board of Education in order to justify its continued use in this battery for the Quality Assurance Program.

Table 15

Summary of Relevance Judgments for Listening Test

	Relevance Judgments *							
Item No.	1	2	3	4	N/R**	Mean	SD	% Crucial/ Important
01	4	6	20	12	1	2.95	.91	76
02	2	8	19	13	1	3.02	.84	76
03	3	11	16	12	1	2.88	.92	67
04	3	7	16	16	1	3.07	.92	76
05	4	7	16	14	2	2.98	.96	73
06	3	6	17	16	1	3.10	.91	79
07	3	9	10	20	1	3.12	.99	71
08	2	7	14	19	1	3.19	.89	79
09	2	7	17	16	1	3.12	.86	79
10	3	8	19	12	1	2.95	.88	74
11	4	5	19	14	1	3.02	.92	79
12	6	7	15	11	4	2.80	1.03	67
13	6	2	16	16	3	3.05	1.04	80
14	7	4	17	12	3	2.85	1.05	73
15	7	7	11	17	1	2.91	1.12	67
16	4	4	13	21	1	3.21	.98	81
17	2	7	17	14	3	3.08	.86	78
18	9	5	13	14	2	2.78	1.15	66
19	7	5	17	12	2	2.83	1.05	71
20	9	10	12	11	1	2.60	1.11	55
21	5	6	19	12	1	2.91	.96	74
22	3	7	14	18	1	3.12	.94	76
23	2	9	11	19	2	3.15	.94	73
24	7	7	13	15	1	2.86	1.10	67
25	5	7	21	9	1	2.81	.92	71
26	3	5	20	14	1	3.07	.87	81
27	5	8	13	16	1	2.95	1.04	69
28	7	7	10	18	1	2.93	1.14	67
29	5	6	18	13	1	2.93	.97	74
30	4	8	18	12	1	2.91	.93	72
31	4	10	14	12	3	2.85	.98	65
32	4	0	8	29	2	3.51	.93	90
33	2	6	10	23	2	3.32	.91	81
34	5	2	13	21	2	3.22	1.01	83
35	2	1	15	23	2	3.44	.78	93
36	3	7	14	17	2	3.10	.94	76
37	2	9	16	14	2	3.02	.88	73
38	2	4	14	21	2	3.32	.85	85
39	1	8	13	19	2	3.22	.85	78
40	3	8	12	17	3	3.08	.97	73
Total Group	164	257	600	634	65	3.03		74

* 4 = crucial, 3 = important, 2 = questionable, 1 = not relevant
** N/R = no response

Table 16
Summary of Relevance Judgments for Writing Test

Item No.	Relevance Judgments *							
	1	2	3	4	N/R**	Mean	SD	% Crucial/ Important
01	2	0	6	41	1	3.75	.66	96
02	1	6	16	27	0	3.38	.78	86
03	3	4	17	20	6	3.23	.89	84
04	4	6	21	19	0	3.10	.91	80
05	4	1	7	38	0	3.58	.88	90
06	1	5	9	35	0	3.56	.76	88
07	2	8	28	11	1	2.98	.75	80
08	2	15	19	13	1	2.88	.86	65
09	2	8	21	13	6	3.02	.82	77
10	2	10	15	23	0	3.18	.89	76
11	2	8	22	17	1	3.10	.82	80
12	5	8	19	10	8	2.81	.94	69
13	3	3	11	33	0	3.48	.86	88
14	4	4	10	32	0	3.40	.95	84
15	1	6	19	24	0	3.32	.77	86
16	1	5	17	26	1	3.39	.76	88
17	0	18	22	10	0	2.84	.74	64
18	1	13	22	14	0	2.98	.80	72
19	2	4	18	25	1	3.35	.81	88
20	4	8	15	23	0	3.14	.97	76
21	3	11	16	19	1	3.04	.94	73
22	3	11	22	13	1	2.92	.86	71
23	3	6	18	22	1	3.20	.89	82
24	5	5	15	15	10	3.00	1.01	75
25	2	7	17	24	0	3.26	.85	82
26	3	8	14	25	0	3.22	.93	78
27	4	11	25	10	0	2.82	.85	70
28	4	5	28	11	2	2.96	.82	81
29	2	7	31	10	0	2.98	.71	82
30	3	7	27	12	1	2.98	.80	80
31	3	6	17	24	0	3.24	.89	82
32	6	11	22	10	1	2.74	.93	65
33	5	4	21	20	0	3.12	.94	82
34	4	4	16	26	0	3.28	.93	84
35	3	17	25	4	1	2.61	.73	59
36	4	16	20	7	3	2.64	.85	57
37	5	14	25	6	0	2.64	.83	62
38	3	7	27	12	1	2.98	.80	80
39	3	12	22	12	1	2.88	.86	69
40	4	13	26	6	1	2.69	.80	65
41	0	8	24	17	1	3.18	.70	84
42	2	17	23	8	0	2.74	.78	62
43	2	6	27	15	0	3.10	.76	84
44	3	8	29	10	0	2.92	.78	78
45	4	12	24	10	0	2.80	.86	68
Total Group	129	373	895	802	51	3.08		77

* 4 = crucial, 3 = important, 2 = questionable, 1 = not relevant

** N/R = no response

Table 17

Summary of Relevance Judgments for Reading Test

	Relevance Judgments *							
Item No.	1	2	3	4	N/R**	Mean	SD	% Crucial/ Important
01	0	4	8	13	1	3.36	.76	84
02	1	0	7	17	1	3.60	.71	96
03	0	2	8	15	1	3.52	.65	92
04	1	3	8	13	1	3.32	.85	84
05	0	2	6	17	1	3.60	.65	92
06	1	2	7	15	1	3.44	.82	88
07	2	3	9	9	3	3.09	.95	78
08	1	2	9	13	1	3.36	.81	88
09	1	3	10	11	1	3.24	.83	84
10	1	6	8	9	2	3.04	.91	71
11	0	3	11	11	1	3.32	.69	88
12	2	8	9	6	1	2.76	.93	60
13	4	3	10	8	1	2.88	1.05	72
14	0	1	9	15	1	3.56	.58	96
15	1	3	8	13	1	3.32	.85	84
16	0	3	8	14	1	3.44	.71	88
17	0	1	4	19	2	3.75	.53	96
18	0	2	5	17	2	3.63	.65	91
19	4	7	11	3	1	2.52	.92	56
20	0	1	10	14	1	3.52	.59	96
21	3	2	9	10	2	3.08	1.02	79
22	1	5	8	11	1	3.16	.90	76
23	2	4	9	10	1	3.08	.94	76
24	3	2	7	11	3	3.13	1.06	78
25	3	2	9	10	2	3.08	1.02	79
26	1	3	12	8	2	3.13	.80	83
27	1	2	9	12	2	3.33	.82	88
28	0	2	7	15	2	3.54	.66	92
29	0	2	8	14	2	3.50	.66	92
30	0	1	9	13	3	3.52	.59	96
Total Group	33	84	252	366	45	3.29		84

* 4 = crucial, 3 = important, 2 = questionable, 1 = not relevant
** N/R = no response

Table 18

Summary of Relevance Judgements for Science Test

	Relevance Judgments *							
Item No.	1	2	3	4	N/R**	Mean	SD	% Crucial/ Important
01	1	13	15	3	0	2.63	.71	56
02	2	11	14	5	0	2.69	.82	59
03	5	14	13	0	0	2.25	.72	41
04	3	11	12	5	1	2.61	.88	55
05	5	9	13	3	2	2.47	.90	53
06	3	2	14	13	0	3.16	.92	84
07	4	12	11	4	1	2.48	.89	48
08	5	12	13	1	1	2.32	.79	45
09	2	11	16	3	0	2.63	.75	59
10	2	5	20	5	0	2.88	.75	78
11	1	3	18	10	0	3.16	.72	88
12	2	7	10	13	0	3.06	.95	72
+13	11	12	5	4	0	2.06	1.01	28
14	1	8	16	7	0	2.91	.78	72
15	3	4	17	8	0	2.94	.88	78
16	5	10	13	4	0	2.50	.92	53
17	4	12	13	3	0	2.47	.84	50
18	4	10	16	2	0	2.50	.80	56
19	4	9	14	5	0	2.63	.91	59
20	4	14	10	2	2	2.33	.80	40
21	5	13	12	2	0	2.34	.83	44
22	5	6	19	2	0	2.56	.84	66
23	4	11	11	5	1	2.55	.93	52
24	7	3	18	4	0	2.59	.98	69
25	2	3	18	8	1	3.03	.80	84
26	7	12	9	3	1	2.26	.93	39
27	3	9	17	2	1	2.58	.77	61
28	2	4	20	6	0	2.94	.76	81
29	5	10	14	3	0	2.47	.88	53
+30	9	10	11	2	0	2.19	.93	41
Total Group	120	270	422	137	11	2.61		59

* 4 = crucial, 3 = important, 2 = questionable, 1 = not relevant
** N/R = no response
+ Items with mean rating below floor of 2.25.

Table 19

Summary of Relevance Judgments for Social Studies Test

	Relevance Judgments *							
Item No.	1	2	3	4	N/R**	Mean	SD	% Crucial/ Important
01	0	7	16	7	1	3.00	.70	76
02	1	3	16	10	1	3.17	.75	87
03	2	4	9	14	2	3.21	.94	79
04	2	6	10	11	2	3.03	.94	72
05	1	6	16	7	1	2.97	.77	77
06	2	8	11	9	1	2.90	.92	67
07	2	8	12	8	1	2.87	.90	67
08	4	9	10	7	1	2.67	.99	57
09	0	9	12	9	1	3.00	.79	70
10	1	3	12	14	1	3.30	.79	87
11	3	4	18	5	1	2.83	.83	77
+12	6	16	8	0	1	2.07	.69	27
13	5	6	9	9	2	2.76	1.09	62
14	2	4	6	18	1	3.33	.96	80
+15	5	14	9	1	2	2.21	.77	34
16	0	5	12	13	1	3.27	.74	83
17	0	4	10	16	1	3.40	.72	77
18	1	4	13	12	1	3.20	.81	83
19	4	11	7	7	2	2.59	1.09	48
20	2	4	15	9	1	3.03	.85	80
21	2	3	16	9	1	3.07	.83	83
22	0	12	11	6	2	2.79	.77	59
23	2	15	10	2	2	2.41	.73	41
24	3	5	15	7	1	2.87	.90	73
25	3	6	13	7	2	2.83	.93	69
26	1	4	16	9	1	3.10	.76	83
27	2	9	12	7	1	2.80	.89	63
28	1	8	8	12	2	3.07	.92	69
29	2	8	11	9	1	2.90	.92	67
30	3	9	11	7	1	2.73	.94	60
Total Group	62	214	354	261	39	2.91		69

* 4 = crucial, 3 = important, 2 = questionable, 1 = not relevant
** N/R = no response
+ Items with mean rating below floor of 2.25.

Table 20

Summary of Relevance Judgments for Literature/Fine Arts Test

	Relevance Judgments *							
Item No.	1	2	3	4	N/R**	Mean	SD	% Crucial/ Important
01	0	11	18	9	0	2.95	.73	71
02	2	12	17	7	0	2.76	.82	63
03	1	5	11	20	1	3.35	.82	84
04	1	4	16	16	1	3.27	.77	86
05	1	3	12	20	2	3.42	.77	89
06	3	20	8	7	0	2.50	.89	40
07	3	13	16	6	0	2.66	.85	58
08	2	6	15	14	1	3.11	.88	78
09	4	13	11	8	2	2.64	.96	53
10	1	14	13	9	1	2.81	.85	59
11	2	5	17	13	1	3.11	.84	81
12	2	2	17	15	2	3.25	.81	89
13	3	4	12	18	1	3.22	.95	81
14	0	10	20	8	0	2.95	.70	74
15	7	15	13	2	1	2.27	.84	41
16	5	9	11	10	3	2.74	1.04	60
17	3	15	15	4	1	2.54	.80	51
18	1	3	16	16	2	3.31	.75	89
19	1	2	15	18	2	3.39	.73	92
20	1	7	15	13	2	3.11	.82	78
21	3	4	16	14	1	3.11	.91	81
22	4	15	8	9	2	2.61	.99	47
23	2	7	19	9	1	2.95	.82	76
24	2	5	17	14	0	3.13	.84	82
25	5	13	12	8	0	2.61	.97	53
26	2	1	16	17	2	3.33	.79	92
27	3	5	14	14	2	3.08	.94	78
28	3	8	16	10	1	2.89	.91	70
29	2	10	14	10	2	2.89	.89	67
30	1	5	12	20	0	3.34	.82	84
31	2	10	12	14	0	3.00	.93	68
32	3	12	15	8	0	2.74	.89	60
33	4	8	15	10	1	2.84	.96	68
34	0	8	11	17	1	3.25	.81	78
35	1	7	18	9	3	3.00	.77	77
Total Group	80	291	503	416	40	2.97		71

* 4 = crucial, 3 = important, 2 = questionable, 1 = not relevant

** N/R = no response

Table 21

Summary of Relevance Judgments for Mathematics Test

	Relevance Judgments *							
Item No.	1	2	3	4	N/R**	Mean	SD	% Crucial/ Important
01	0	1	6	18	0	3.68	.56	96
02	1	0	16	8	0	3.24	.66	96
03	0	2	9	13	1	3.46	.66	92
04	3	3	11	8	0	2.96	.98	76
05	5	12	7	1	0	3.08	.72	79
06	1	8	8	8	0	2.92	.91	64
+07	7	9	7	1	1	2.08	.88	33
08	0	3	10	11	1	3.33	.70	87
09	3	4	12	6	0	2.84	.94	72
10	1	6	8	9	1	3.04	.91	71
11	1	7	9	8	0	2.96	.89	68
12	1	2	8	14	0	3.40	.82	88
13	1	3	12	9	0	3.16	.80	84
14	1	6	14	3	1	2.79	.72	71
15	2	6	13	3	1	2.71	.81	67
16	0	4	11	9	1	3.21	.72	83
17	2	11	7	5	0	2.60	.91	48
18	1	7	11	6	0	2.88	.83	68
19	2	7	10	6	0	2.80	.91	64
20	0	5	5	14	1	3.38	.82	79
21	0	6	10	9	0	3.12	.78	76
+22	5	11	6	2	1	2.21	.88	33
+23	7	10	5	0	3	1.91	.75	23
24	3	5	11	6	0	2.80	.96	68
25	0	4	9	11	1	3.29	.75	83
Total Group	42	135	240	194	14	2.96		71

* 4 = crucial, 3 = important, 2 = questionable, 1 = not relevant
** N/R = no response
+ Items with mean rating below floor of 2.25.

Table 22

Panelists Responses to Two Questions Concerning the Content Validity and Relevance of the Essay Test

Question 1: Please circle your response to the statement, "The NTE Essay Examination which I have reviewed, including its administration and scoring procedures, is a satisfactory measure of general writing skill."

Frequency of Response					Summary Statistics	
(1) Strongly Agree	(2) Agree	(3) ?	(4) Disagree	(5) Strongly Disagree	Mean	SD
7(12%)	37(62%)	10(17%)	3(5%)	3(5%)	2.30	.93

Question 2: Please consider the writing skills measured by the NTE Essay Test and provide us with your judgments of how important those skills are to success in the teacher education program with which you are most familiar. Specifically, we would like to know how important it would be for a student to have minimal writing skills, as measured by the NTE Essay test.

Frequency of Response				Summary Statistics	
(4) Crucial	(3) Important	(2) Questionable Importance	(1) Irrelevant	Mean	SD
33(55%)	26(43%)	1(2%)	0(0%)	3.53	.54

Table 23

Summary of Content Relevance Ratings for Total Panel

	Number of Panelists	Number of Items	Inter-rater Reliability	Percent of Relevance* Ratings by Response Choice N 1	Q 2	IMP 3	CRU 4	OMIT	Percent Responses Marked Important or Crucial	Percent Items Marked Important/Crucial by Majority of Raters	Percent Items With Average Score Greater or Equal to 2.25
Communication Skills				b					c	d	a
1. Listening	43	40	.58	10	16	36	38	4	74	100	100
2. Writing	50	45	.68	6	17	41	36	2	77	100	100
3. Reading	26	30	.72	4	11	34	50	6	84	100	100
General Knowledge											
4. Science	32	30	.76	16	27	43	14	1	59	70	93
5. Social Studies	31	30	.72	7	24	40	29	4	69	87	93
6. Literature/FA	38	35	.79	6	23	39	32	3	71	91	100
7. Mathematics	25	25	.88	7	22	39	32	2	71	84	88

*Percentages for 1-4 based on total valid responses; percentage for omit based on total number of panelists.

Table 23a

Aggregate Indices for the Evaluation of Relevance Ratings

Index and Description	Evaluation of the Index
a) % of items with mean ≥ 2.25; the average score for each item is calculated on a 1-4 scale and reported in Tables 15-21; the percentage of items on the test with averages equal or greater than 2.25 is then calculated.	- satis. (+): 90-100% - marg. (?): 85-89% - unsat. (-): 0-84%
b) % of relevance ratings "Not Relevant"; the total number of "not relevant" ratings is divided by the total number of valid ratings and a percent is calculated.	- satis. (+): 0-10% - marg. (?): 11-12% - unsat. (-): 13-100%
c) % of ratings "Crucial"/"Important"; total number of "crucial" and "important" ratings is divided by the total number of valid ratings and a percent is calculated.	- satis. (+): 75-100% - marg. (?): 70-74% - unsat. (-): 0-69%
d) % of items where majority of raters responded "Crucial"/"Important"; the number of items where the percentage of "crucial" and "important" items is equal to or greater than 51% is calculated; that number is divided by the total number of items and a percentage is calculated.	- satis. (+): 75-100% - marg. (?): 70-74% - unsat. (-): 0-69%
e) average of all responses on 1-4 scale; the average score for the total test is calculated on a 1-4 scale and is reported in Tables 15-21.	- satis. (+): 2.25-4.0 - marg. (?): 2.20-2.24 - unsat.(-): 1.00-2.19

Table 24
Comparison of Relevance Judgments of College University Faculty and Public School Faculty: Selected Summary Variables

Test	Group	N*	Percent of Relevance Ratings By Response Choice**					Percent Responses Marked Important/ Crucial	Percent Items Marked Important/Crucial by Majority of Raters	Percent Items With Average Equal to or Greater Than 2.25	Weighted Total Average***
			Not 1	Ques 2	Imp 3	Cru 4	Omit				
Communication Skills											
Listening	C-U	25	13	15	35	37	06	72	100	100	2.96
	PS	18	06	16	38	40	01	78	100	100	3.09
Writing	C-U	36	06	15	41	38	03	79	100	100	3.19
	PS	14	04	22	41	33	01	74	93	98	3.07
Reading	C-U	21	05	13	34	48	07	82	100	100	3.28
	PS	05	01	07	35	57	02	92	100	100	3.66
General Knowledge											
Science	C-U	23	12	27	46	15	00	61	80	87	2.60
	PS	09	14	32	41	13	03	54	63	70	2.45
Soc. Stu.	C-U	23	07	24	40	29	04	69	90	93	2.88
	PS	08	06	24	39	31	05	70	90	90	2.73
Lit/FA	C-U	24	06	21	37	36	04	73	91	100	3.06
	PS	14	07	25	42	26	04	68	86	94	2.89
Math	C-U	17	09	23	38	30	03	68	84	88	2.81
	PS	08	02	21	42	35	01	77	88	100	3.18

* Total number of judges available.

** Percent of responses 1-4 based on all non-omitted responses; percent of omit based on total number of available judges.

*** Average of valid non-missing responses on all relevance items; computed only in cases for which there were no missing cases.

Table 25

Panelists Responses to Two Questions Concerning the Relevance of the Essay Test
Responses by University/College Faculty and by Public School Teachers

Question 1: Please circle your response to the statement, "The NTE Essay examination which I have reviewed, including its administration and scoring procedures, is a satisfactory measure of general writing skill."

	Frequency of Response					Summary Statistics	
	(1) Strongly Agree	(2) Agree	(3) ?	(4) Disagree	(5) Strongly Disagree	Mean	SD
University/ College	5(14%)	21(58%)	5(14%)	3(8%)	2(6%)	2.33	1.01
Public School	2(8%)	16(67%)	5(21%)	0(0%)	1(4%)	2.25	.79

Question 2: Please consider the writing skills measured by the NTE Essay test and provide us with your judgments of how important those skills are to success in the teacher education program with which you are most familiar. Specifically, we would like to know how important it would be for a student to have minimal writing skills, as measured by the NTE Essay test.

	Frequency of Response				Summary Statistics	
	(4) Crucial	(3) Important	(2) Questionable Importance	(1) Irrelevant	Mean	SD
University/ College	22(61%)	13(36%)	1(3%)	0(0%)	3.58	.55
Public School	11(46%)	13(54%)	0(0%)	0(0%)	3.46	.51

Instructional Validity Data

Communications Skills Test. The results of the instructional validity judgments made on individual items for the Reading, Writing and Listening multiple choice subtests are summarized in Table 26. The number of "yes" answers to the question which asked whether at least 90% of students applying for admission to teacher education would have had a reasonable opportunity to receive instruction and the percentage of valid "yes" responses are reported for each question. The number and percentage of "no" responses are also reported. Table 28 summarizes the number of "yes" responses to individual items across individuals on the subtest. Various descriptive statistics are presented. The same table presents descriptive statistics on the four point Global Scale. Table 29 also summarizes the frequency of responses to the four-point Global Scale, for the Communication Skills subtests. The frequency and percentage of "yes" responses given in answer to the instructional validity question for each blueprint category are given in Table 29. The responses to the two instructional validity questions for the Essay Test are summarized in Table 30. Table 31 summarizes aggregate data for the multiple-choice Communication Skills subtests. Table 32 evaluates that evidence as "satisfactory", "marginally satisfactory" or "unsatisfactory". It should be observed that the judgment standard given to the participants, ie., that 90% or more of their students would have had a reasonable opportunity to be instructed, is fairly conservative.

Taken as a whole, the Communication Skills Test received satisfactory instructional validity ratings. This was especially true for the Reading and Writing multiple-choice subtests. Tables 31 and 32 summarize the results of the analysis for six instructional validity indices. All of the indices for both of those tests were judged satisfactory. For example, all items on both tests were answered affirmatively by a majority of raters. On average, eighty-seven percent of the items on the Reading subtest were answered "yes"; the comparable figure for the Writing multiple-choice subtest was 86%. The other evidence for the Reading and Writing subtests is similar. It should be pointed out that one might evaluate each of the indices for instructional validity quite differently. For example, index #4, the percentage of blueprint categories where the percentage "yes" is equal to or greater than 75%, is a very crude index. On a test such as Reading with only three blueprint categories, the difference between two of three, and three of three on the index is a substantially large percentage. Such differences in indices need to be accounted for during interpretation and analysis of data.

The Listening subtest analysis resulted in four indices rated as "acceptable" and two as "marginally acceptable". The marginal indexes were: index (#3) computed on the Global rating scale, fell just below the standard for the lowest satisfactory value (51%). The evidence from the document analysis on the Listening subtest was mixed, ranging from low to marginal.

The Essay subtest instructional validity evidence was based on two questions. Since the instructional validity evidence was in fact based on only two data points one must be somewhat cautious with respect to its interpretation. The first question asks whether at least 90% of the students applying for admission to teacher education would have had a reasonable opportunity for instruction in the skills which are described in the ETS description of Writing Skills for the Essay subtest. Of thirty-seven college/university participants, fifteen "strongly agreed", and sixteen "agreed". Only five "disagreed". We feel that this is satisfactory evidence in support of the instructional validity of the Essay Test. The second item asked the same question but asked the participants to determine whether 90% of students would have a reasonable opportunity to be instructed in the skills and knowledge _actually_ measured by the test. Of a total of thirty-six participants, 12 "strongly agreed" and 14 "agreed". Only five "disagreed" or "strongly disagreed". This evidence is not as strong as that for the first question but is at least marginally acceptable.

In summary, we feel that most students applying for admission to the teacher education programs in North Carolina would have had a reasonable opportunity to receive instruction in the skills and knowledge assessed by the Communication Skills subtests in Reading, Writing Multiple Choice, Listening and Essay-Writing.

General Knowledge Test. The instructional validity judgments made on individual items for the Science, Social Studies, Mathematics and Literature/Fine Arts subtests are summarized in Table 27. Statistics describing the average percent of items answered "yes" for the instructional validity are reported in

Table 28; additional descriptive statistics and frequency distributions summarizing results for the Global item are presented in Tables 28 and 29. Table 29 also includes summaries of the frequency and percentage of "yes" responses for each category of the test blueprint. Table 30 reports on the results for the two items used to assess the instructional validity of the Essay subtest.

Table 31 summarizes all the individual evidence for the subtests into six indices which are defined in Table 31a. It may be seen that in general the instructional validity data for the General Knowledge Tests is somewhat more mixed than that of the Communication Skills Test. The percentage of items where a majority marked "yes" for each of these tests is less than 100% which was consistently the percentage for the Communication Skills subtest. For example, 77% of the Science Test items and 92% of the Mathematics Test items are so marked. The average percentage of test items marked "yes" by participants ranged from 68% (Science) to 86% (Mathematics). The average of the percentage of "yes" responses across blueprint categories (index #5) varied more than the corresponding percentage for the Communication Skills subtests; the lowest in Social Studies (66%) and the highest in Science (88%).

The same set of rules were defined, which were used to evaluate each of the Communication Skills subtests indices as "satisfactory", "marginally satisfactory" or "unsatisfactory". The results of those evaluations are given in Table 32.

The instructional validity evidence for the Mathematics subtest was satisfactory. Ninety-two percent of the Mathematics subtest items were answered "yes". The average percentage of "yes" items on the test was 86. Over fifty percent of the answers to the Global question were in the "very similar" or "no appreciable difference" category. Two thirds of the blueprint categories had 75% or more "yes" responses and the average percentage of "yes" responses was 81. The document analysis ratings were generally high.

The evidence for the Literature/Fine Arts and Science subtests is less definitive. In both cases the number of "satisfactory" indices are greater than the number of "unsatisfactory". The Science subtest has three of the six indices evaluated as "satisfactory", one "marginally satisfactory" and two as "unsatisfactory". One of the negative indices is the average percentage of items marked "yes" at 68%. This value just falls below the lower bound of the evaluati/n criterion (by 2% points). The Literature/Fine Arts subtest had two satisfactory indices and three which were "marginally satisfactory". Only one was "unsatisfactory" and that is index #4, which is a relatively crude index; if just one more of the three blueprint categories was greater than 75%, then the index would have been evaluated as "satisfactory". In general, we conclude that the evidence supports the instructional validity of both the Science and Literature/Fine Arts subtests.

The evidence for the Social Studies subtest is not nearly as strong. This subtest produced one "satisfactory" index and three "marginally satisfactory". Two were "unsatisfactory". If the "marginally acceptable" and "acceptable" indices are combined, they outweigh the "unacceptable" indices. Evidence based on individual item judgments (indices #1 and #2 on Table 31) tend to be satisfactory. For example, 87% of the items on the Social Studies subtest were marked "yes" and the average percent "yes" answers on the subtest was 73%. Respondents to the Global item which is by definition more general, and perhaps less useful, tended to respond that there were "some differences" (63%) between the test content and the curriculum at their institution. The Social Studies subtest blueprint ratings which were more specific evaluations, generally were low to marginally satisfactory.

When asked in the Evaluation Questionnaire, given at the conclusion of data gathering, about their confidence in the NTE, approximately 50% of the Social Studies subtest participants expressed confidence. This was somewhat less than the percentage for the total set of judges (58%). When asked about their degree of confidence about their instructional validity judgments, there was very little difference between the Social Studies participants and the total group of participants. Some participants in their written comments, suggested that much of this subtest consisted of the interpretation of figures and charts rather than questions on the substance of Social Studies, and observed that Reading Comprehension Skills would help the students do well. As was the case with the Science subtest, discussed earlier, this subtest differs from the Mathematics and Literature/Fine Arts subtests in its complexity. The test assesses many of the Social Sciences (eg. Sociology, Psychology, History, Geography, Political Science etc.). As such, it is difficult to construct a test which reflects all of the different Social Studies curricula at the various institutions in the state. Furthermore, even given the different curricula represented, there are undoubtedly large differences in individual student programs in Social Studies within a given institution. Because of this one might expect the ratings for this test to be somewhat lower. However we feel that the evidence in support of the instructional validity of the Social Studies subtest is only barely marginal. While the evidence is not clearly negative, neither is it positive.

We conclude that students applying for admission to teacher education programs in the State will have had a reasonable opportunity to receive instruction in preparation for the subtests in Reading, multiple-choice Writing, Essay, Science, Mathematics and Literature/Fine Arts. The evidence is less clear for the Social Studies subtest, but suggests that the test is barely marginally acceptable. We recommend that the State might begin to use the General Knowledge and Communication Skills Test but that it should continue to study the instructional validity of the Social Studies subtest.

Table 26

Instructional Validity Judgments by College/University Panelists on Individual Items: Communication Skills Subtest (Reading, Writing, Listening)

	Reading Test (N = 21)*		Writing Test (N = 36)		Listening Test (N = 25)	
Item No.	Yes (%)	NR (%)**	Yes (%)	NR (%)	Yes (%)	NR (%)
01	18(90)	1(05)	33(100)	3(08)	20(91)	3(12)
02	20(95)	0(00)	29(94)	5(14)	19(83)	2(08)
03	18(90)	1(05)	28(97)	7(19)	19(83)	2(08)
04	17(85)	1(05)	23(82)	8(22)	17(81)	4(16)
05	17(85)	1(05)	32(100)	4(11)	19(83)	2(08)
06	18(86)	0(00)	33(100)	3(08)	16(76)	4(16)
07	13(72)	3(14)	29(91)	4(11)	20(91)	3(12)
08	16(89)	3(04)	25(81)	5(14)	21(96)	3(12)
09	18(95)	2(10)	29(97)	6(17)	20(87)	2(08)
10	16(89)	3(14)	20(69)	7(19)	15(68)	3(12)
11	18(90)	1(05)	31(94)	3(08)	18(82)	3(12)
12	13(68)	2(10)	23(92)	11(31)	19(91)	4(16)
13	15(83)	3(14)	29(91)	4(11)	16(76)	4(16)
14	20(100)	1(05)	30(94)	4(11)	14(70)	5(20)
15	20(95)	0(00)	32(97)	3(08)	15(68)	3(12)
16	21(100)	0(00)	30(91)	3(08)	19(86)	3(12)
17	18(86)	0(00)	26(84)	5(14)	20(95)	4(16)
18	19(91)	0(00)	22(71)	5(14)	17(74)	2(08)
19	10(71)	7(33)	31(94)	3(08)	12(60)	5(20)
20	17(90)	2(10)	22(79)	8(22)	20(91)	3(12)
21	16(84)	2(10)	22(76)	7(19)	18(82)	3(12)
22	14(82)	4(19)	23(77)	6(17)	20(91)	3(12)
23	10(59)	4(19)	31(97)	4(11)	21(96)	3(12)
24	16(94)	4(19)	25(93)	9(25)	17(81)	4(16)
25	14(78)	3(14)	28(90)	5(14)	19(91)	4(16)
26	15(83)	3(14)	27(87)	5(14)	21(91)	2(08)
27	19(95)	1(05)	26(81)	4(11)	20(87)	2(08)
28	17(90)	2(10)	28(93)	6(17)	19(86)	3(12)
29	14(74)	2(10)	26(84)	5(14)	16(80)	5(20)
30	18(95)	2(10)	26(87)	6(17)	15(71)	4(16)
31			29(91)	4(11)	15(71)	4(16)
32			22(79)	8(22)	19(83)	2(08)
33			25(83)	6(17)	20(87)	2(08)
34			27(90)	6(17)	18(86)	4(16)
35			22(73)	6(17)	16(80)	5(20)

Table 26 (cont.)

	Reading Test (N = 21)*		Writing Test (N = 36)		Listening Test (N = 25)	
Item No.	Yes (%)	NR (%)**	Yes (%)	NR (%)	Yes (%)	NR (%)
36			19(70)	9(25)	14(70)	5(20)
37			16(62)	10(28)	15(75)	5(20)
38			22(73)	6(17)	21(91)	2(08)
39			20(71)	8(22)	20(87)	2(08)
40			19(70)	9(25)	21(91)	2(08)
41			26(90)	7(19)		
42			23(77)	6(17)		
43			28(90)	5(14)		
44			24(83)	7(19)		
45			24(86)	8(22)		

* Total number of panelists for instructional validity judgments.

** Percentage of yes responses based on total "yes" or "no" responses; percentage of no response, responses based on total number of panelists.

Table 27

Instructional Validity Judgments by College/University Panelists on Individual Items: General Knowledge Subtest (Mathematics, Science, Social Studies, Literature/Fine Arts)

	SCIENCE (N=23)*		SOCIAL STUDIES (N=23)		LITERATURE/FA (N=24)		MATHEMATICS (N=17)	
ITEM	YES(%)	**NR(%)	YES(%)	NR(%)	YES(%)	NR(%)	YES(%)	NR(%)
01	18(86)	2(09)	19(86)	1(04)	16(76)	3(13)	15(100)	2(12)
02	19(83)	0(00)	21(96)	1(04)	12(75)	8(33)	17(100)	0(00)
03	10(50)	3(13)	11(69)	7(30)	13(72)	6(25)	16(94)	1(06)
04	13(59)	1(04)	16(76)	2(09)	17(85)	4(17)	15(88)	2(12)
05	16(70)	0(00)	10(50)	3(13)	17(85)	4(17)	12(75)	1(06)
06	19(86)	1(04)	9(50)	5(22)	11(58)	5(21)	14(88)	1(06)
07	18(78)	0(00)	14(70)	3(13)	11(50)	2(08)	4(29)	3(18)
08	9(45)	3(13)	16(80)	3(13)	12(63)	5(21)	16(94)	0(00)
09	20(91)	1(04)	14(74)	4(17)	15(75)	4(17)	15(88)	0(00)
10	18(82)	1(04)	19(86)	1(04)	17(77)	2(08)	15(88)	0(00)
11	22(96)	0(00)	18(90)	3(13)	15(79)	5(21)	15(88)	0(00)
12	13(59)	1(04)	5(26)	4(17)	12(67)	6(25)	17(100)	0(00)
13	9(47)	4(17)	10(53)	4(17)	17(85)	4(17)	16(100)	1(06)
14	13(62)	2(09)	18(86)	2(09)	18(86)	3(13)	15(94)	1(06)
15	15(71)	2(09)	6(33)	5(22)	9(45)	4(17)	12(80)	2(12)
16	10(48)	2(09)	18(82)	1(04)	9(50)	6(25)	16(94)	0(00)
17	10(46)	1(04)	16(80)	3(13)	13(65)	4(17)	11(69)	1(06)
18	10(48)	2(09)	16(84)	4(17)	14(74)	5(21)	13(77)	0(00)
19	17(81)	2(09)	13(65)	3(13)	16(84)	5(21)	15(94)	1(06)
20	13(62)	2(09)	18(86)	2(09)	12(71)	7(29)	16(94)	0(00)
21	11(61)	5(22)	16(80)	3(13)	15(71)	3(13)	14(88)	1(06)
22	13(65)	3(13)	10(63)	7(30)	10(53)	5(21)	13(81)	1(06)
23	10(56)	5(22)	10(59)	6(26)	15(79)	5(21)	0(00)	7(41)
24	16(73)	1(04)	15(75)	3(13)	16(76)	3(13)	14(88)	1(06)
25	15(68)	1(04)	14(74)	4(17)	9(50)	6(25)	16(100)	1(06)
26	9(47)	4(17)	16(80)	3(13)	13(68)	5(21)		
27	12(60)	3(13)	16(73)	1(04)	12(63)	5(21)		
28	22(96)	0(00)	13(65)	3(13)	13(65)	4(17)		
29	17(77)	0(00)	13(77)	6(26)	9(50)	6(25)		
30	11(69)	7(30)	14(78)	5(22)	20(87)	1(04)		
31					20(87)	1(04)		
32					17(81)	2(08)		
33					11(58)	5(21)		
34					17(81)	3(13)		
35					10(53)	5(21)		

*Total number of panelists

**Percentage "Yes" is based on total valid (excluding no judgment) responses.

Percentage "NR" (no response) is based on total number of judges making response

Table 28

Descriptive Statistics for Instructional Validity Indices: Percent of Items on Test Marked "Yes" and Global Rating Item

	COMMUNICATION SKILLS SUBTEST			GENERAL KNOWLEDGE SUBTEST			
	LIS	WRI	REA	SCI	SOC	LIT	MAT
Percent Items "YES" For Individuals							
Mean	82.52	85.94	86.57	68.22	72.96	72.78	85.82
SD	20.78	19.50	18.73	23.83	25.59	28.09	16.66
Minimum	30	12	40	20	17	0	48
Maximum	100	100	100	100	100	100	100
N*	23	33	21	23	22	23	17
Global Rating**							
Mean	2.54	1.57	1.95	2.52	2.59	2.52	2.47
SD	.93	.71	.69	.51	.59	.67	.62
Minimum	1	1	1	2	1	2	2
Maximum	4	3	3	3	3	4	4
N*	24	33	20	23	22	23	17

**"Very Similar" scored as 1; "Not Appreciably Different" as 2; "Some Appreciable Differences" as 3; "Little Similarity as 4.

*Number of valid responses.

Table 29

Frequency Distribution* for Global Rating and Blueprint Items: Instructional Validity

		COMMUNICATION SKILLS			GENERAL KNOWLEDGE			
		LIS	WRI	REA	SCI	SOC	LIT	MAT
Global Rating								
Very Similar	A	3(13%)	18(55%)	5(25%)	0 (0%)	1 (5%)	0 (0%)	0 (0%)
Not Appreciably	B	9(37%)	11(33%)	11(55%)	11(48%)	7(32%)	13(56%)	10(59%)
Some Difference	C	8(33%)	4(12%)	4(20%)	12(52%)	14(63%)	8(35%)	6(35%)
Little Similarity	D	4(17%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	2 (9%)	1 (6%)
No Response/Omit		1 (4%)	0 (0%)	1 (0%)	0 (0%)	1 (4%)	1 (4%)	0 (0%)
		Frequency and Percent of "YES" Responses to Blueprint Category						
Blueprint Items**								
1		21(88%)	32(100%)	21(100%)	21 (91%)	17(77%)	17(74%)	17(100%)
2		18(75%)	31 (97%)	18 (86%)	23(100%)	12(57%)	17(77%)	8 (47%)
3		19(79%)	26 (81%)	17 (81%)	20 (87%)	17(77%)	16(70%)	16 (94%)
4		20(83%)	25 (81%)		23(100%)	11(52%)		16 (94%)
5			29 (94%)		18 (78%)			10 (63%)
6			32(100%)		19 (83%)			15 (88%)
7			30 (94%)		18 (78%)			
8			32(100%)		18 (78%)			
9			31(100%)		22 (96%)			
10			29(91%)					
11			29(91%)					
12			32(100%)					

*Percentages are based on valid responses only; "NR" = No Response are based on total number of panelists.

**Number of Blueprint sections vary by test; see section for description of each.

Table 30

Panelists Responses to Two Questions Concerning the Instructional Validity of the Essay Test

Question 1: Please circle your response to the statement, "At least 90% of the students applying for admission to teacher education at the institution with which I am most familiar have had a reasonable opportunity to receive instruction in the writing skills described in the ETS statement before applying for admission to teacher education."

Frequency of Response					Summary Statistics	
(1) Strongly Agree	(2) Agree	(3) ?	(4) Disagree	(5) Strongly Disagree	Mean	SD
15(41%)	16(43%)	1(3%)	5(14%)	0(0%)	1.89	.99

Question 2: Please circle your response to the statement, "At least 90% of the students applying for admission to teacher education at the institution with which I am most familiar have had a reasonable opportunity to receive instruction in the writing skills that are actually measured by the NTE Essay Test."

Frequency of Response					Summary Statistics	
(1) Strongly Agree	(2) Agree	(3) ?	(4) Disagree	(5) Strongly Disagree	Mean	SD
12(33%)	14(39%)	5(14%)	4(11%)	1(3%)	2.11	1.09

Table 31

Aggregate Statistics Summarizing Instructional Validity Evidence (a)

Test	% Items Majority Marked "YES"	Average % of "YES" on Items	Global: % "Very Similar"/"No Appreciable Diff."	Blueprint: % of Categories Where % "YES" ≥ 75% (b)	Blueprint: Average of % "YES" in Categories for Test	Document Analysis (c)
	1	2	3	4	5	6
Com. Skills						
Reading	100	87	80	3/3(100)	89	M/H
Writing	100	86	88	12/12(100)	94	H
Listening	100	83	50	4/4(100)	81	L/M
Gen. Know.						
Science	77	68	48	9/9(100)	88	L
Soc. Studies	87	73	37	2/4 (50)	66	M
Lit/Fine Art	86	73	56	1/3 (33)	74	M
Mathematics	92	86	59	4/6 (66)	81	H

(a) See Table 31a for description of the indices and the criteria for flagging a test for further analysis.

(b) Note that Index #5 is very imprecise since the number of categories in the Blueprint can be as few as three.

(c) See Vol. 3 of Phase II Report (June 30, 1983) for detailed description of results; scale indicates likelihood that topic is covered in the courses taken:
L = Low, M = Moderate, H = High.

Table 31a

Criteria for Evaluating Instructional Validity Ratings

Index	Evaluation of Index
A. Item Based Questions	
1. Percent for which majority say "yes" to "90%" question; the percentage of valid responses marked "yes" for each item is calculated and reported in Tables 26-27 the number of items for which the percentage is equal to, or greater than 51% is divided by the total number of items and a percentage is calculated and reported in Table 31.	- satis. (+): 75-100% - marg. (?): 70-74% - unsat. (-): 1-69%
2. Percent of items marked "yes" to "90%" question, averaged across individuals; the percentage of items on the test marked "yes" by each participant is calculated; the aver- of the percentages is reported in Tables 28 and 31.	- satis. (+): 75-100% - marg. (?): 70-74% - unsat. (-): 0-69%
B. Global Scale	
3. Percent in categories "very similar" /"not appreciable"; the total number of responses to the Global Question which were "very similar" or "no appreciable differences" were calculated and reported in Table 29. That number was divided by the total number of valid responses and a percentage was calculated and reported in Table 31.	- satis. (+): 51-100% - marg. (?): 46-50% - unsat. (-): 0-45%
C. Blueprint Judgements	
4. Percent of categories where percent of "yes" was ≥ 75%; the percentage of "yes" responses to the "90%" question was calculated for each blueprint category and reported in Table 29. The number of blueprint categories for which the percentage was equal to or greater than 75% was divided by the number of categories and a percentage calculated.	- satis. (+): 51-100% - marg. (?): 46-50% - unsat. (-): 0-45%

Table 31a cont.

Index	Evaluation of the Index
5. Average of percent "yes" in categories; the percentages of "yes" responses for each blueprint category are averaged for each test and reported in Table 31.	- satis. (+): 75-100% - marg. (?): 70-74% - unsat. (-): 0-69%
D. Document Analysis	
6. Rating: L (small likehood topic covered in course) M (moderate likehood topic covered in course) H (high likehood topic covered in course) The weighted means reported in the Document Analysis for each blueprint category were inspected and classified as: Low (ratings less than 1.75, Moderate (ratings 1.75 to 2.25), High (greater than 2.25)	- satis. (+): High (H), M/H - marg. (?): Moderate (M), L/M - unsat. (-): Low (L)

Table 32

Summary of Indicators[a] : Relevance and Instructional Validity

	Relevance Index					Instructional Validity Index					
Test	a	b	c	d	e	1	2	3	4	5	6
Listening	+	+	?	+	+	+	+	?	+	+	?
Writing	+	+	+	+	+	+	+	+	+	+	+
Reading	+	+	+	+	+	+	+	+	+	+	+
Science	+	-	-	?	+	+	-	?	+	+	-
Social Studies	+	+	-	+	+	+	?	-	?	-	?
Literature/FA	+	+	?	+	+	+	?	+	-	?	?
Mathematics	?	+	?	+	+	+	+	+	+	+	+
Essay	+	+				+	?				

Key "+" = Satisfactory; "?" = Borderline, Marginally Satisfactory;
"-" = Unsatisfactory, Low

The essay test relevance ratings consisted of two questions. Instructional validity ratings consisted of two questions.

(a) Key to indices found in Tables 23a and 31a.

Results for Individual Subtests

Separate panels of participants recommended standards for each of the subtests that compose the NTE General Knowledge and Communication Skills Tests. Two separate judgment sessions were used to elicit panelist's recommendations on the seven multiple-choice subtests. Three judgment sessions were used for the Essay subtest.

Distributions of the standards recommended for each of these subtests were summarized by computing a variety of descriptive statistics. Methods used to analyze the data are described in Section II, above. The results of these analyses are shown in Tables 33 and 34, for the first and second judgment sessions, respectively. The results shown in Table 34 for the Essay subtest were derived from the third, rather than the second, judgment session. Because final recommended standards were based on data collected during the second judgment for the multiple-choice tests and data collected during the third judgment session for the Essay subtest, it seemed appropriate to display summaries of these data in a common table.

The results of separate analyses of recommendations provided by public school teachers and college/university faculty are shown, for the second judgment session, in Tables 35 and 36. Once again, summaries of recommendations on appropriate standards for the Essay subtest, based on data from the third judgment session, are displayed in these tables.

From the data in Tables 33 through 36 it is clear that participants° recommendations on appropriate standards for these

Table 33

Summary statistics for distributions of raw-score standards recommended for individual subtests of the General Knowledge and Communication Skills Tests of the National Teacher Examinations by all participants (first judgment session).

Subtest	Minimum	P_{25}	P_{50}	P_{75}	Maximum	Mean
Social Studies	4.38	10.34	13.34	16.04	21.04	13.78
Mathematics	7.38	8.61	10.41	12.31	15.71	10.75
Literature/ Fine Arts	6.59	12.02	14.73	17.53	28.52	15.55
Science	5.42	9.23	11.63	12.64	14.64	11.13
Listening	11.50	14.90	19.90	22.29	32.30	19.46
Reading	6.10	11.05	14.06	17.15	26.75	15.07
Writing (multiple-choice)	8.19	16.33	20.14	23.98	29.68	20.38
Writing Essay	3.00	5.11	6.63	7.65	10.50	6.27

Table 34

Summary statistics for distributions of raw-score standards recommended for individual subtests of the General Knowledge and Communication Skills Tests of the National Teacher Examinations by all participants (second judgment session).

Subtest	Minimum	P_{25}	P_{50}	P_{75}	Maximum	Mean
Social Studies	10.34	12.24	13.84	15.22	19.84	14.05
Mathematics	7.11	8.41	9.41	10.71	15.01	9.80
Literature/ Fine Arts	6.67	14.03	15.73	17.92	26.33	16.36
Science	7.52	10.04	12.04	13.34	15.61	11.78
Listening	7.46	16.45	19.49	21.10	27.60	19.29
Reading	7.42	11.00	13.65	16.35	20.89	14.07
Writing (multiple-choice)	12.91	18.48	20.57	22.68	29.17	20.72
Writing Essay	3.00	5.24	6.01	6.70	10.00	5.94

Table 35

Summary statistics for distributions of raw-score standards recommended for individual subtests of the General Knowledge and Communication Skills Tests of the National Teacher Examinations by public school teachers (second judgment session).

Subtest	Minimum	P_{25}	P_{50}	P_{75}	Maximum	Mean
Social Studies	12.34	12.64	14.44	15.84	18.22	15.02
Mathematics	7.11	7.61	9.11	10.71	11.41	9.53
Literature/ Fine Arts	10.10	14.03	15.73	17.52	22.53	16.27
Science	7.52	8.44	12.74	13.74	15.61	12.02
Listening	16.46	19.40	20.60	21.26	26.70	20.84
Reading	13.34	13.35	14.65	16.55	16.65	15.30
Writing (multiple-choice)	18.37	18.84	19.48	21.27	23.98	20.26
Writing Essay	4.50	6.09	6.91	7.79	8.50	6.81

Table 36

Summary statistics for distributions of raw-score standards recommended for individual subtests of the General Knowledge and Communication Skills Tests of the National Teacher Examinations by college and university faculty (second judgment session).

Subtest	Minimum	P_{25}	P_{50}	P_{75}	Maximum	Mean
Social Studies	10.34	11.64	13.14	14.24	19.84	13.69
Mathematics	7.11	8.41	9.40	10.21	15.01	9.92
Literature/ Fine Arts	6.67	13.32	15.63	18.23	26.33	16.41
Science	8.31	10.04	12.04	12.94	15.54	11.70
Listening	7.46	15.49	18.10	20.50	27.60	18.13
Reading	7.42	10.85	13.15	15.15	20.89	13.84
Writing (multiple-choice)	12.91	16.98	20.88	23.68	29.18	20.90
Writing Essay	4.00	5.61	6.67	7.79	10.00	6.51

subtests varied moderately to substantially. However, much of the variation in participants° recommendations occurred in the extremes of the distributions; the interquartile ranges were generally small. As will be shown in later tables, the standard errors of the mean recommendations on these subtests were quite small, indicating that participant samples of adequate size were used to secure recommendations.

Differences between the means of standards recommended by public school teachers and college/university faculty were generally modest. The means of public school teachers° reocmmendations were higher than those of college/university faculty for five of the eight subtests. The reverse was true for the other three. The data in Tables 35 and 36 therefore support an interpretation of essential consistency between the recommendations of both types of participants.

Internal consistency reliabilities (Cronbach°s alpha values), standard errors of measurement, standard errors of mean recommended standards, and "total" standard errors were computed for each subtest, using data from the final judgment sessions. Details of methods used to derive these statistics are provided in Section II. Results are shown in Table 37.

From the data in Table 37, it is clear that the major component of random error in establishing recommended test standards can be attributed to the unreliability of the subtests rather than participant sampling. For every subtest, the standard error of measurement is far larger than the standard error of the mean recommended test standard. Since the squares of these standard errors are summed when computing a "total" standard error, random error due to the unreliability of the tests is the dominant error factor.

Results for the General Knowledge and Communication Skills Tests

Recommended standards for the General Knowledge and Communication Skills Tests of the NTE were computed by combining the standards recommended for the individual subtests that compose these tests. A linear combination of standards recommended on the Social Studies, Mathematics, Literature and Fine Arts, and Science subtests was formed to compute recommended standards for the General Knowledge Test. A linear combination of standards recommended for the Listening, Reading, Writing multiple-choice and Essay subtests was formed to compute recommended standards for the Communication Skills Test. The equations that define these linear combinations are shown in Section II, above. Linear transformations provided by Educational Testing Service, also shown in Section II, were used to place the recommended standards on the score scales ETS uses in reporting examinees° performances on the two tests.

Recommended standards for the General Knowledge and Communication Skills tests were analyzed and displayed in a variety of ways. Table 38 contains descriptive statistics for the distributions of standards recommended by public school teachers, college and university faculty, and all participants for both of these tests, during the final judgment session. This table is structurally similar to Tables 33 through 36, shown above.

Table 37

Internal consistency reliability coefficients, estimated standard errors of measurement, standard errors of mean recommended standards, and "total" standard errors for the eight subtests that compose the General Knowledge and Communication Skills Tests of the National Teacher Examinations. Estimates of reliability and standard errors of measurement are based on data resulting from the April, 1983 administration of the NTE in North Carolina.

Subtest	Internal Consistency Reliability	Standard Error of Measurement	Standard Error of the Mean Recommended Standard	Total Standard Error
Social Studies	0.8171	2.4197	0.4285	2.4574
Mathematics	0.8535	2.0991	0.3817	2.1335
Literature/ Fine Arts	0.8580	2.5494	0.6754	2.6373
Science	0.8263	2.4122	0.4041	2.4458
Listening	0.8561	2.6440	0.6363	2.7195
Reading	0.8545	2.3524	0.6968	2.4534
Writing (multiple-choice)	0.8781	2.8466	0.4969	2.8896
Essay	0.7999	0.9753	0.1690	0.9898

Table 38

Summary statistics for distributions of scaled-score standards recommended by public school teachers, college and university faculty, and all participants, for the General Knowledge and Communication Skills Tests of the National Teacher Examinations during the final judgment session.

Recommendations on the General Knowledge Test

Type of Participant	Minimum	P25	P50	P75	Maximum	Mean
Public School Teachers	628	632	639	643	650	639
College and University Faculty	624	632	637	641	657	639
All Participants	624	633	638	643	657	639

Recommendations on the Communication Skills Test

Type of Participant	Minimum	P25	P50	P75	Maximum	Mean
Public School Teachers	637	642	645	649	652	646
College and University Faculty	622	635	642	648	663	643
All Participants	622	636	643	648	663	643

It might at first appear logical to recommend that the standards for the General Knowledge and Communication Skills Tests be set at some measure of central tendency (either the mean or the median) of the distribution of participants° recommended scaled-score standards. In fact, we find the median to be more conceptually appealing than the mean because it is less sensitive to outlying recommendations and, for these data, the loss in precision due to use of the median rather than the mean is of trivial consequence. This follows since the major component of random error is clearly the unreliability of the tests rather than errors due to random sampling of participants (see Table 37, above).

However, if the median were to be used as the recommended standard, 50 percent of the applicants to teacher education programs who had true scores at the recommended standard would fail each test, solely due to random (mostly measurement) errors. A false-negative error rate of this magnitude is intollerable. We therefore computed median recommended standards on each test, adjusted downward by one, two, and three "total" standard errors. In addition, we computed the false-negative error rates associated with each of these standards. The results of these computations are shown in Tables 39 through 44 for the second of the error models discussed in Section II (the one based on total-test screening of applicants). Tables 39 and 40 contain recommended standards for the General Knowledge and Communication Skills Tests, respectively, based on the judgments of all participants. Tables 41 and 42 contain recommendations, based solely on the judgments of college and university faculty.

Table 39

Standards for the General Knowledge Test of the National Teacher Examinations, as Recommended by Public School Teachers and College/University Faculty in North Carolina (All Participants)

Reference Point	Scaled Score	Percent Failing*	Percent Failing by Chance Error**
Median	638	30.4	50.0
Median Minus One Std. Error	634	24.1	15.9
Median Minus Two Std. Errors	631	19.7	2.3
Median Minus Three Std. Errors	627	11.6	0.0

* Percent of North Carolina teacher education applicants and students who were tested in April, 1983 who would have failed the test, had the passing score been set at the indicated scaled score.

** Percent of North Carolina teacher education applicants and students with true scaled scores at the median recommended passing score who would fail the test due to errors of measurement, because their total General Knowledge Test observed scaled scores would be lower than the scaled score corresponding to the indicated reference point.

Table 40

Standards for the Communication Skills Test of the National Teacher Examinations, as Recommended by Public School Teachers and College/University Faculty in North Carolina (All Participants)

Reference Point	Scaled Score	Percent Failing*	Percent Failing by Chance Error**
Median	643	24.0	50.0
Median Minus One Std. Error	639	18.1	15.9
Median Minus Two Std. Errors	636	14.5	2.3
Median Minus Three Std. Errors	632	8.8	0.0

* Percent of North Carolina teacher education applicants and students who were tested in April, 1983 who would have failed the test, had the passing score been set at the indicated scaled score.

** Percent of North Carolina teacher education applicants and students with true scaled scores at the median recommended passing score who would fail the test due to errors of measurement, because their total Communication Skills Test observed scaled scores would be lower than the scaled score corresponding to the indicated reference point.

Table 41

Standards for the General Knowledge Test of the National Teacher Examinations, as Recommended by College and University Faculty in North Carolina

Reference Point	Scaled Score	Percent Failing*	Percent Failing by Chance Error**
Median	637	29.0	50.0
Median Minus One Std. Error	634	24.1	15.9
Median Minus Two Std. Errors	630	18.0	2.3
Median Minus Three Std. Errors	626	11.1	0.0

* Percent of North Carolina teacher education applicants and students who were tested in April, 1983 who would have failed the test, had the passing score been set at the indicated scaled score.

** Percent of North Carolina teacher education applicants and students with true scaled scores at the median recommended passing score who would fail the test due to errors of measurement, because their total General Knowledge Test observed scaled scores would be lower than the scaled score corresponding to the indicated reference point.

Table 42

Standards for the Communication Skills Test of the National Teacher Examinations, as Recommended by College and University Faculty in North Carolina

Reference Point	Scaled Score	Percent Failing*	Percent Failing by Chance Error**
Median	642	22.6	50.0
Median Minus One Std. Error	639	18.1	15.9
Median Minus Two Std. Errors	635	12.4	2.3
Median Minus Three Std. Errors	631	8.0	0.0

* Percent of North Carolina teacher education applicants and students who were tested in April, 1983 who would have failed the test, had the passing score been set at the indicated scaled score.

** Percent of North Carolina teacher education applicants and students with true scaled scores at the median recommended passing score who would fail the test due to errors of measurement, because their total Communication Skills Test observed scaled scores would be lower than the scaled score corresponding to the indicated reference point.

Table 43

Standards for the General Knowledge Test of the National Teacher Examinations, as Recommended by Public School Teachers in North Carolina

Reference Point	Scaled Score	Percent Failing*	Percent Failing by Chance Error**
Median	639	32.5	50.0
Median Minus One Std. Error	635	25.5	15.9
Median Minus Two Std. Errors	631	19.7	2.3
Median Minus Three Std. Errors	627	11.6	0.0

* Percent of North Carolina teacher education applicants and students who were tested in April, 1983 who would have failed the test, had the passing score been set at the indicated scaled score.

** Percent of North Carolina teacher education applicants and students with true scaled scores at the median recommended passing score who would fail the test due to errors of measurement, because their total General Knowledge Test observed scaled scores would be lower than the scaled score corresponding to the indicated reference point.

Table 44

Standards for the Communication Skills Test of the National Teacher Examinations, as Recommended by Public School Teachers in North Carolina

Reference Point	Scaled Score	Percent Failing*	Percent Failing by Chance Error**
Median	645	28.3	50.0
Median Minus One Std. Error	642	22.6	15.9
Median Minus Two Std. Errors	638	16.1	2.3
Median Minus Three Std. Errors	635	12.4	0.0

---

* Percent of North Carolina teacher education applicants and students who were tested in April, 1983 who would have failed the test, had the passing score been set at the indicated scaled score.

** Percent of North Carolina teacher education applicants and students with true scaled scores at the median recommended passing score who would fail the test due to errors of measurement, because their total Communication Skills Test observed scaled scores would be lower than the scaled score corresponding to the indicated reference point.

Tables 43 and 44 display recommendations based on the judgments of public school teachers.

Any use of tests for screening applicants for employment or admission to education programs has associated risks of false-negative and false-positive errors. In the application proposed by the State Board of Education in North Carolina, false-negative errors occur when an able candidate for admission to a teacher education program is denied admission due to measurement errors or random errors resulting from the selection of the sample of participants used to recommend test standards. The burden of false-negative errors therefore falls primarily upon the applicant, and upon the state only to the extent that it might be denied the services of an able classroom teacher. Conversely, the burden of false-positive errors falls primarily on the state. Admission of an unable applicant for a teacher education program, again solely because of errors of measurement or random sampling errors associated with the selection of participants used to recommend test standards, is costly to the state and society. To the extent that admissions tests are increasingly reliable, both types of errors are reduced. However, in the face of imperfectly reliable tests and finite samples of participants in standard-setting studies, it behooves the state to err on the side of the applicant. Indeed, legally-based procedural due process considerations permit no alternative.

The results shown in Tables 39 through 44 clearly illustrate the unacceptably high risks of false-negative errors associated with adoption of median recommended standards. Adoption of a standard that is one "total" standard error below the median reduces the risk of a false-negative error for an applicant with ability equal to the median, from 50 percent to just under 16 percent. Even a risk of this magnitude might be deemed procedurally inequitable. Adoption of a standard that is two total standard errors below the median reduces the risk of a false-negative error for an applicant with true ability at the median to a negligible 2.3 percent; a figure that would seem to be acceptable.

In addition to consideration of false-negative error rates, the resonableness of any recommended standard must be examined in terms of its consequences. Although it is impossible to predict with certainty, the failure rates that will result from use of the NTE General Knowledge and Communication Skills Tests to screen applicants to North Carolina°s teacher education programs, it is possible to make some tentative suggestions on the basis of data collected in the April, 1983 administration of the NTE in North Carolina. It must be realized that these data were collected very late in the academic year, from a volunteer sample of North Carolina students who were not at risk of being denied admission to a teacher education program regardless of their test performances. Thus the percentages of those students who would have failed the tests, had the recommended standards been imposed, might not be useful predictors of the failure rates that will be realized in an operational testing program. Since the consequences of failing the NTE tests will be severe, we would expect students who take the tests in the future to be highly motivated. We would also expect higher education institutions with approved teacher education programs to exert some considerable effort to see that applicants to their programs are well prepared to take the NTE tests. These factors suggest operational failure rates that are somewhat lower than those realized in the April, 1983 test administration. However, to the degree that the April, 1983 volunteer sample of examinees was biased in favor of more able students (a possibility that cannot be examined with the data at hand), we would expect the failure rates realized in the April, 1983 test administration to be conservative.

Table 45 contains data on the percentage of April, 1983 examinees who would have failed at least one of the General Knowledge or Communication Skills Tests, had various recommended standards been employed. The data presented in this table show that the slightly more lenient standards recommended by college and university faculty would have resulted in passing four to five percent more applicants than the standards recommended by public school teachers. Had the median standards recommended by all participants been adopted, almost four out of ten applicants would have failed at least one of the tests. Had the median recommended standard been reduced by two "total" standard errors, just under one out of four applicants would have failed at least one of the tests. Ensuring that fewer than one percent of the applicants with true abilities equal to the median recommended standards would be denied admission, solely because of random errors, would have resulted in a failure rate just under 17 percent.

Table 45

Percent of April, 1983 North Carolina Teacher Education Applicants and Students Who Would Have Failed One or Both of the General Knowledge and Communication Skills Tests of the National Teacher Examinations, Given Various Recommended Passing Scores.

	Type of Participant		
Reference Point	Public School Teachers	College and University Faculty	All Participants
Median	39.2	35.0	37.0
Median minus 1 S.E.	33.2	29.8	29.8
Median minus 2 S.E.	25.2	22.7	24.3
Median minus 3 S.E.	19.0	14.9	16.9

In Section II we described an alternative model for protecting applicants from risks of false-negative errors. This model presumed that applicants to teacher education programs would be selected on the basis of their performances on each of the _subtests_ that compose the General Knowledge and Communication Skills Tests. Were that procedure to be adopted, it would be necessary to ensure that the risk of false-negative errors was acceptably small _for each subtest_.

The risk of false-negative errors could be controlled on a subtest-by-subtest basis by reducing the median raw-score recommended standard for each subtest, by multiples of the raw-score "total" standard error for that subtest. Resulting adjusted raw-score recommended standards would then be combined through the weighting equations used by Educational Testing Service to determine raw scores on the General Knowledge and Communication Skills Tests. These raw scores would then be converted to scaled scores through the linear transformations provided by Educational Testing Service.

These procedures were applied to the recommendations resulting from the final judgment session for each subtest. Recommended standards are contained in Tables 46 through 51. Median recommendations provided by all participants, adjusted by one, two, and three "total" standard errors on each subtest, are shown in Tables 46 and 47, for the General Knowledge and Communication Skills Tests, respectively. Similarly-derived statistics, based on the recommendations of college and university faculty, are shown in Tables 48 and 49. Finally, Tables 50 and 51 contain statistics on the recommendations of public school teachers.

If the data in Tables 46 through 51 are compared with corresponding values in Tables 39 through 44, the effect of protecting applicants from risks of false-negative errors, on a subtest-by-subtest basis, becomes clear. The resulting standards are substantially more lenient than those derived from the total-test error model. In fact, adjustment by two "total" standard errors using the total-test error model has the same result as adjustment by one "total" standard error using the subtest-by-subtest error model.

Since current State Board of Education policy would screen applicants to North Carolina teacher education programs on the basis of their total scores on the General Knowledge and Communication Skills Tests, the total-test error model would seem to be more appropriate than the subtest-by-subtest model. If, at some future date, the Board°s policy is modified so that applicants must demonstrate some minimum level of competence on each of the subtests that compose these tests, the results shown in Tables 46 through 51 should be applied.

Table 46

Standards for the General Knowledge Test of the National Teacher Examinations, as Recommended by Public School Teachers and College/University Faculty in North Carolina (All Participants).

Reference Point	Scaled Score	Percent Failing*	Percent Failing by Chance Error**
Median	638	30.4	50.0
Median Minus One Std. Error	631	19.7	15.9
Median Minus Two Std. Errors	624	7.0	2.3
Median Minus Three Std. Errors	616	0.5	0.0

* Percent of North Carolina teacher education applicants and students who were tested in April, 1983 who would have failed the test, had the passing score been set at the indicated scaled score.

** Percent of North Carolina teacher education applicants and students with true scaled scores at the median recommended passing score who would fail each of the subtests in the NTE General Knowledge Test due to errors of measurement, because their observed scaled scores would be lower than the scaled score corresponding to the indicated reference point.

Table 47

Standards for the Communication Skills Test of the National Teacher Examinations, as Recommended by Public School Teachers and College/University Faculty in North Carolina (All Participants)

Reference Point	Scaled Score	Percent Failing*	Percent Failing by Chance Error**
Median	643	25.7	50.0
Median Minus One Std. Error	636	14.9	15.9
Median Minus Two Std. Errors	629	6.6	2.3
Median Minus Three Std. Errors	623	1.9	0.0

* Percent of North Carolina teacher education applicants and students who were tested in April, 1983 who would have failed the test, had the passing score been set at the indicated scaled score.

** Percent of North Carolina teacher education applicants and students with true scaled scores at the median recommended passing score who would fail each of the subtests in the NTE Communication Skills Test due to errors of measurement, because their observed scaled scores would be lower than the scaled score corresponding to the indicated reference point.

Table 48

Standards for the General Knowledge Test of the National Teacher Examinations, as Recommended by College and University Faculty in North Carolina

Reference Point	Scaled Score	Percent Failing*	Percent Failing by Chance Error**
Median	637	29.1	50.0
Median Minus One Std. Error	630	18.0	15.9
Median Minus Two Std. Errors	623	5.2	2.3
Median Minus Three Std. Errors	615	0.3	0.0

* Percent of North Carolina teacher education applicants and students who were tested in April, 1983 who would have failed the test, had the passing score been set at the indicated scaled score.

** Percent of North Carolina teacher education applicants and students with true scaled scores at the median recommended passing score who would fail each of the subtests in the NTE General Knowledge Test due to errors of measurement, because their observed scaled scores would be lower than the scaled score corresponding to the indicated reference point.

Table 49

Standards for the Communication Skills Test of the National Teacher Examinations, as Recommended by College and University Faculty in North Carolina

Reference Point	Scaled Score	Percent Failing*	Percent Failing by Chance Error**
Median	642	22.6	50.0
Median Minus One Std. Error	635	12.4	15.9
Median Minus Two Std. Errors	629	6.6	2.3
Median Minus Three Std. Errors	622	1.0	0.0

* Percent of North Carolina teacher education applicants and students who were tested in April, 1983 who would have failed the test, had the passing score been set at the indicated scaled score.

** Percent of North Carolina teacher education applicants and students with true scaled scores at the median recommended passing score who would fail each of the subtests in the NTE Communication Skills Test due to errors of measurement, because their total Communication Skills Test observed scaled scores would be lower than the scaled score corresponding to the indicated reference point.

Table 50

Standards for the General Knowledge Test of the National Teacher Examinations, as Recommended by Public School Teachers in North Carolina

Reference Point	Scaled Score	Percent Failing*	Percent Failing by Chance Error**
Median	639	32.5	50.0
Median Minus One Std. Error	631	19.7	15.9
Median Minus Two Std. Errors	624	7.0	2.3
Median Minus Three Std. Errors	616	0.5	0.0

* Percent of North Carolina teacher education applicants and students who were tested in April, 1983 who would have failed the test, had the passing score been set at the indicated scaled score.

** Percent of North Carolina teacher education applicants and students with true scaled scores at the median recommended passing score who would fail each of the subtests in the NTE General Knowledge Test due to errors of measurement, because their observed scaled scores would be lower than the scaled score corresponding to the indicated reference point.

Table 51

Standards for the Communication Skills Test of the National Teacher Examinations, as Recommended by Public School Teachers in North Carolina

Reference Point	Scaled Score	Percent Failing*	Percent Failing by Chance Error**
Median	645	28.4	50.0
Median Minus One Std. Error	639	18.1	15.9
Median Minus Two Std. Errors	632	8.8	2.3
Median Minus Three Std. Errors	625	2.6	0.0

* Percent of North Carolina teacher education applicants and students who were tested in April, 1983 who would have failed the test, had the passing score been set at the indicated scaled score.

** Percent of North Carolina teacher education applicants and students with true scaled scores at the median recommended passing score who would fail each of the subtests in the NTE Communication Skills Test due to errors of measurement, because their observed scaled scores would be lower than the scaled score corresponding to the indicated reference point.

CONCLUSIONS AND RECOMMENDATIONS

Use of the NTE in North Carolina

The relevance of all four subtests of the Communication Skills Test was judged to be acceptable. The skills and knowledge assessed by two of the four subtests of the General Knowledge Test -- the Social Studies and the Literature Fine Arts subtests -- were judged to be clearly relevant to success in North Carolina's Teacher Education programs; evidence in support of the relevance of the Mathematics subtest was judged to be satisfactory but was somewhat weaker than that of the Social Studies and Literature Fine Arts subtests. The Science subtest was found to be marginally relevant, and borders on being classified as unacceptable on this validity dimension.

The instructional validity of the Communication Skills test was judged to be satisfactory. Judgments on the instructional validity of the General Knowledge Test were mixed. The Mathematics, Literature and Fine Arts, and Science subtests were all found to be at least acceptable with respect to instructional validity. The Social Studies subtest was judged to be barely marginal in its instructional validity.

Results of the validity analyses described in the preceeding sections, considered collectively, support a decision to initiate use of the NTE General Knowledge and Communication Skills Tests in selecting students for North Carolina's teacher education programs. However, it is essential that the predictive validity of these tests be investigated as soon as possible. The need for additional validity evidence is especially critical for the Science subtest and the Social Studies subtest, since evidence on the validity of these subtests barely supports their initial use in selecting students for North Carolina's teacher education programs.

Recommended Test Standards

Based on the results of the procedures described in an earlier section of this report, we recommend that the passing score on the General Knowledge Test be 631 points and that the passing score on the Communication Skills Test be 636 points, on the scale used by Educational Testing Service to report scores on these tests. Applicants would have to earn scores that were equal to or higher than these values, to gain admission to approved teacher education programs in North Carolina.

The performance of North Carolina students in the April 1983 administration of the National Teacher Examinations allows us to examine the consequences of these recommended passing scores. The results are as follows: almost 20 percent of North Carolina examinees would have failed the General Knowledge Test and over 14 percent of North Carolina examinees would have failed the Communication Skills Test, had these passing scores been used. More than 24 percent of North Carolina examinees would have failed at least one of these tests.

Essential considerations and actions

Adoption of the recommendations contained in this report by the State Board of Education and the Board of Governors of the University of North Carolina will have substantial impact on applicants to the state's teacher education programs, and on the institutions that provide those programs. How these students and programs will be affected will depend, in large part, on the policies these Boards adopt for the operation of an NTE screening program.

Although the development of policies for program operation is outside the focus of this study, we would like to alert the Boards to several essential issues that require their timely consideration.

North Carolina institutions that provide teacher education programs and students considering careers in teaching must be informed of the Boards' policy decisions on admission to teacher education at the earliest possible date. Although these institutions are aware of the state's intent to use the National Teacher Examinations in selecting students for teacher education programs, they are not yet aware of the required passing scores, the dates of administration of the NTE, and applicants' opportunities for retesting. Since provision of adequate notice has been found by the courts to be a major consideration in constitutional due process litigation, we suggest that the Boards seek the counsel of the State Attorney General's Office prior to the initiation of the NTE screening program.

The Boards must consider policies on applicants' opportunities to repeat the National Teacher Examinations, should they fail in their initial attempts. Relevant issues include (1) the number of times applicants will be allowed to repeat the examinations, (2) a minimum time interval between repeated attempts to pass the examinations, (3) whether an examinee who completes one examination successfully but fails the other (e.g., an examinee who passes the General Knowledge Test but fails the Communication Skills Test) will be required to repeat both examinations or only the one failed, (4) whether examinees who repeat an examination will be required to take an alternate form of the examination or will be allowed to repeat an identical form (The availability of alternative forms of the examinations from Educational Testing Service may constrain this policy). We suggest that the Boards also seek the counsel of the State Attorney General's Office on these issues. Previous court cases involving the use of testing for selection have focused on procedural due process considerations that include provision of adequate remedy to those who fail to be admitted. Remedies that have been deemed appropriate include opportunities for retesting and institutional provision of opportunities to remediate deficiencies in knowledge and ability. Thus the Boards might consider policies on remediation programs for applicants who cannot pass the National Teacher Examinations.

Many applicants to teacher education programs in North Carolina complete their first two years of higher education in the state's community colleges. It is therefore essential that policies on dissemination of information regarding the NTE screening program include consideration of the community colleges and their students. The State Board of Education and the Board of Governors of the University of North Carolina might consider seeking the endorsement of the State Board for the North Carolina Community Colleges on policies regarding use of the NTE in screening applicants to the state's teacher education programs.

Present Educational Testing Service policies limit the dissemination of scores on the National Teacher Examinations to agencies and institutions requested by individual examinees. These policies might limit the ability of North Carolina institutions to provide timely information to applicants on the status of their admission to teacher education programs. Alternative policies that facilitate the operation of the NTE screening program might have to be negotiated with Educational Testing Service.

Although the results of this study support the initial use of the National Teacher Examinations in screening applicants to North Carolina's teacher education programs, continued use of these tests demands an ongoing program of review and research. In particular, the state must study the differential impact of an NTE screening program on various legally protected groups, such as racial and ethnic minorities, women, etc. Although it is clear that the Boards have no discriminatory intent in adopting policies that limit admission to teacher education programs to applicants who earn passing scores on the NTE General Knowledge and Communication Skills Tests, it is possible that the policies will have unintended differential effects on legally protected groups. Should these differential effects be found, the reasonableness of the Boards' policies might be subject to legal scrutiny.

We strongly recommend that the State Board of Education initiate a study of the relationship between applicants' scores on the NTE General Knowledge and Communication Skills Tests, and their actual performance in teacher education programs. Such a study is essential to continued use of the NTE tests for screening applicants to teacher education programs. Although the courts have found that validation of tests through expert judgment is sufficient for initial use of the tests in screening programs, continued use requires empirical demonstrations of validity, as would be obtained in the study proposed here.

In addition to satisfying legal requirements for continued use of the NTE tests in screening teacher education applicants, the proposed study would provide an evaluation of the costs and benefits of using the NTE in addition to, or in lieu of, applicants' scores on other readily-available predictors of performance in teacher education programs. Other predictors of performance, such as high school grade-point average, high school rank, first-year college grade-point average, scores on the Scholastic Aptitude Tests (SAT), and grade-point averages in selected college subjects, might prove to be adjuncts to, or substitutes for, NTE scores, in selecting able students for teacher education programs.

Additionally, a predictive validity study of the NTE would provide essential evidence on the validity of the Science and Social Studies subtests in screening teacher education applicants. Since these two subtestsprograms.

Additionally, a predictive validity study of the NTE would provide essential evidence on the validity of the Science and Social Studies subtests in screening teacher education applicants. Since these two subtests were found to be marginal in the validity assessments reported in this study, their continued use in a battery of screening tests can only be justified through a demonstration of their usefulness in predicting students' success in teacher education programs.

The Boards' concurrent resolution establishing the Quality Assurance Program required that higher education institutions with approved teacher education programs counsel prospective students on the likelihood of their success in passing the NTE tests and their likely success in teacher education. A well-designed predictive validity study of the NTE Core Batteries I and II would provide these institutions with information necessary to offer students such guidance.

Finally, we suggest that the Boards commission continuing evaluations of the impact of the NTE screening program on the quality of public education in North Carolina. Specifically, we recommend that the Boards monitor such variables as: the number of applicants to teacher education programs, the average SAT scores of applicants to teacher education programs, the student retention and graduation rates of teacher education programs; the relationship between students' performances on Core Batteries I and II of the NTE prior to admission to teacher education, and their performances on Core Battery III of the NTE following completion of teacher education programs; and the availability of certified teachers in various certification categories. Ultimately, the quantity and quality of professional personnel available to, and functioning in, the State's classrooms should be assessed.